

# Computational Mechanisms for Action Selection

Toby Tyrrell

PhD  
University of Edinburgh  
1993



## Declaration

I declare that this thesis has been composed by myself and that the research reported herein has been conducted by myself unless otherwise indicated.

Some of Part I has appeared in the paper

Tyrrell, T. and Mayhew, J.E.W. 1991. Computer Simulation of an Animal Environment. In *Proceedings of the First International Conference on Simulation of Adaptive Behavior*. Meyer, J-A. and Wilson, S.W. (eds). MIT Press/Bradford Books.

The arguments in section 7.2 have appeared as

Tyrrell, T. 1992. Defining the Action Selection Problem. In the *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum Associates.

Some of chapter 11 is to be published as

Tyrrell, T. 1993. The Use of Hierarchies for Action Selection. In the *Journal of Adaptive Behavior*, 1:4, 1993, to appear. Also in the *Proceedings of the Second International Conference on Simulation of Adaptive Behavior*. Meyer, J-A., Wilson, S.W. and Roitblat, H. (eds). MIT Press/Bradford Books, 1993.

Edinburgh, 19th March, 1993.

Toby Tyrrell.



## Abstract

Imagine a zebra in the African savannah. At each moment in time this zebra has to weigh up alternative courses of action before deciding which will be most beneficial to it. For instance, it may want to graze because it is short of food, or it may want to head towards a water hole because it is short of water, or it may want to remain motionless in order to avoid detection by the predator it can see lurking nearby. This is an example of the problem of *action selection*: how to choose, at each moment in time, the most appropriate out of a repertoire of possible actions.

This thesis investigates action selection in a novel way and makes three main contributions. Firstly, a description is given of a simulated environment which is an extensive and detailed simulation of the problem of action selection for animals. Secondly, this simulated environment is used to investigate the adequacy of several theories of action selection such as the drive model, Lorenz's hydraulic model and Maes' spreading activation network. Thirdly, a new approach to action selection is developed which determines the most appropriate action in a principled way, and which does not suffer from the inherent shortcomings found in other methods.

## Acknowledgements

I would like to thank both of my supervisors, John Hallam and David Willshaw, for their ever-willing donations of time and advice. I owe a lot to David, in particular, for providing such a congenial and supportive working environment and for the many occasions on which he has gone out of his way to help me. It has been much appreciated. Thanks also to Rosanna Maccagnano for helping me out with various onerous duties (and for many issues of PN!), and to Jay Buckingham and Andrew Gillies for providing such excellent and reliable computing facilities.

Phil MacLauchlan, Neil Thacker and Patrick Courtney were good friends at Sheffield and their company made the time there much more enjoyable. In Edinburgh, Peter Dayan, Martin Simmen, Carl Rasmussen, Marcus Frean, Jay Buckingham, Geoff Goodhill, Andrew Gillies and Nick Taylor have all been very sociable and stimulating colleagues; thanks for the numerous valuable discussions, cakes and chocolate cookies.

Many thanks to Martin Simmen, Bridget Hallam, Peter Dayan, Lisa Meeden, Janet Halperin, David Willshaw and John Hallam who have all helped me greatly by providing very useful comments on drafts of this thesis.

My parents have assisted me financially and have provided a lot of encouragement, for which I am very grateful. Finally, very heartfelt thanks to Helen Pitman for her companionship, her support, and her patience, which have all made completing this thesis that much easier.

I am grateful for support from the SERC (grant no. 89310818).

# Contents

<b>Prologue</b>	<b>1</b>
Terminology . . . . .	3
Overview . . . . .	5
<b>I A Simulated Environment</b>	<b>10</b>
<b>1 Introduction</b>	<b>11</b>
1.1 Testing Theories of Action Selection Using Robots . . . . .	14
1.2 The Value of Computer Simulations . . . . .	16
1.3 Other Work on Simulated Environments . . . . .	17
1.4 Plan of Part I . . . . .	24

<b>2</b>	<b>Theory</b>	<b>25</b>
2.1	Prior Assumptions . . . . .	25
2.2	Complexity and Realism . . . . .	26
2.3	Genetic Fitness . . . . .	26
2.4	Sub-Problems . . . . .	29
2.5	Modelling Perception, Navigation and Motor Control . . . . .	32
<b>3</b>	<b>Description</b>	<b>35</b>
3.1	Animal's Health . . . . .	39
3.2	Implementation of Sub-Problems . . . . .	41
3.3	Perception . . . . .	51
3.4	Navigation . . . . .	53
3.5	Motor Control . . . . .	62
3.6	Analytical Tools . . . . .	63
<b>4</b>	<b>Validation</b>	<b>70</b>
4.1	Low-Level Tests . . . . .	71
4.2	High-Level Tests . . . . .	73
4.3	Obtaining Statistically Significant Results . . . . .	92

<b>5 Discussion</b>	<b>95</b>
5.1 Assumptions . . . . .	95
5.2 Using the Simulated Environment . . . . .	98
5.3 The Fallibility of Results from the Simulated Environment . . . . .	100
5.4 Terminology . . . . .	101
5.5 Theoretically Optimal Actions . . . . .	103
<b>6 Conclusions to Part I</b>	<b>106</b>
6.1 Further Work . . . . .	107
<b>II Evaluation of Action Selection Mechanisms</b>	<b>109</b>
<b>7 Introduction</b>	<b>110</b>
7.1 Scope of Testing Procedure . . . . .	110
7.2 A Taxonomy for Sub-Problems . . . . .	112
7.3 Other Terminology . . . . .	117
7.4 Plan of Part II . . . . .	118

<b>8</b>	<b>Mechanisms for Action Selection</b>	<b>120</b>
8.1	Drives . . . . .	121
8.2	Tinbergen . . . . .	124
8.3	Lorenz . . . . .	128
8.4	Baerends . . . . .	132
8.5	Maes . . . . .	133
8.6	Rosenblatt & Payton . . . . .	142
8.7	Halperin . . . . .	145
8.8	Other Related Work . . . . .	148
8.9	Summary of Differences Between Mechanisms . . . . .	157
<b>9</b>	<b>Performance of Mechanisms</b>	<b>160</b>
9.1	How Each Mechanism was Tested . . . . .	160
9.2	Factors Affecting the Desirability of Different Alternatives . . . . .	163
9.3	Implementational Issues and Performance Results . . . . .	168

<b>10 Theoretical Analysis</b>	<b>209</b>
10.1 Dealing with All Types of Sub-Problem . . . . .	209
10.2 Compromise Candidates . . . . .	210
10.3 Proper Treatment of Information . . . . .	211
10.4 Combination of Stimuli . . . . .	212
10.5 Requirements for an Action Selection Mechanism . . . . .	213
10.6 Summary of Findings on Each ASM . . . . .	216
10.7 Summary of Testing . . . . .	219
<b>11 How Best to Compute Action Selection</b>	<b>223</b>
11.1 Theory Behind the New Mechanism . . . . .	224
11.2 Examples . . . . .	242
11.3 Discussion . . . . .	244
<b>12 Conclusions to Part II</b>	<b>251</b>
12.1 Further Work . . . . .	252
<b>Epilogue</b>	<b>254</b>
<b>Appendix A: Details of the Implementation of the Extended Rosenblatt &amp; Payton Mechanism</b>	<b>256</b>

# Prologue

This thesis is an investigation of both the problem of *action selection* and the computational mechanisms for providing solutions to that problem.

The problem of action selection is that of choosing at each moment in time the most appropriate action out of a repertoire of possible actions (e.g. drinking, sleeping, moving, grooming). The problem can also be thought of as one of time-allocation; the animal has to decide how to apportion its available time so as to simultaneously satisfy many needs. Action selection is separate from perception (calculation and classification of stimuli) and motor control (execution of actions). The process of making a decision as to which action is most appropriate will need to take into account many stimuli, some of which will be external (e.g. nearby predators, food sources, mates) and some of which will be internal (e.g. blood sugar levels, body temperature). An example situation in which several different stimuli are important is shown in figure 0.1. Different action selection mechanisms take these stimuli and compute a preferred action in different ways. The simulated environment to be described in the first part of this thesis provides a method for testing and comparing these different computational approaches to action selection.

The action selection mechanisms considered here are somewhat similar to planning systems, in that they produce sequences of actions which achieve useful objectives. However, the mechanisms considered here are more reactive (responding to the immediate state of the environment) than 'deliberative' (following long-term plans formulated



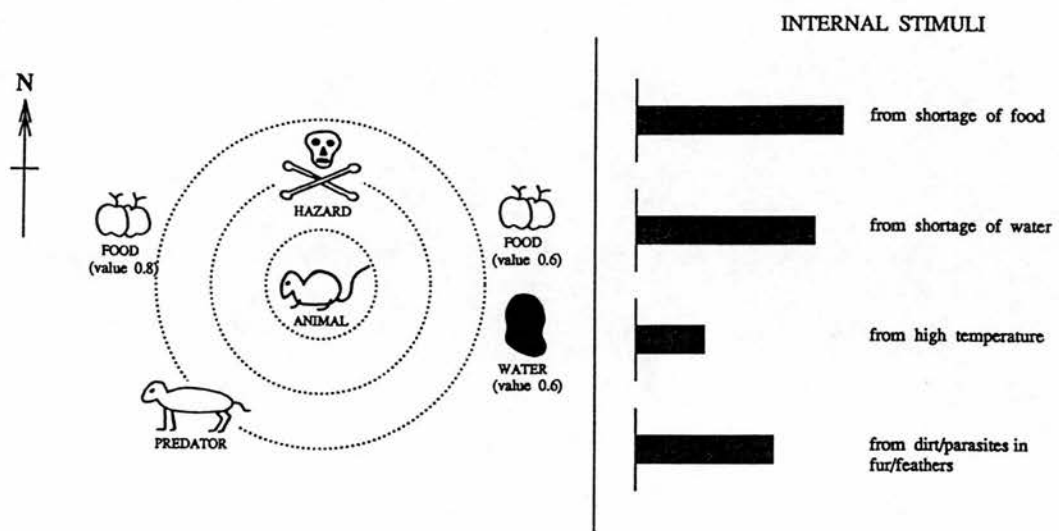


Figure 0.1: An example situation in which many different stimuli need to be taken into account in order to decide what the best action is. In this case the animal has several feasible actions to choose from: should it move south directly away from the hazard, move north-east directly away from the predator, move west towards the high-valued food source, move east towards the average-valued food and the average-valued water source, or stay where it is and clean itself? The stimuli shown here are offsets from optimal values, and the length of each bar is proportional to the seriousness of the offset.

some time in the past). It is worth noting that while the mechanisms considered here do not involve explicit plans, they are to some extent *goal-driven*, in that the behaviour will tend to be oriented towards correcting disturbances in homeostatic variables or rectifying undesirable external situations. They are, in the main, designed to account for how animals can produce appropriate behaviour in natural environments, rather than to explain how people can solve traditional planning problems such as how to make coffee or how to stack blocks on top of one another in a certain way.

The mechanisms considered here do *not* incorporate learning. The aim here is to examine the sorts of algorithms or computational methods which can be used to

select actions, rather than to examine the ways in which the selection of actions can be improved by experience. For this reason, methods such as temporal difference learning [Sutton, 88] and Q-learning [Watkins, 89] are not covered in any depth here.

## Terminology

Some terms which will be used extensively throughout this thesis are now defined in conjunction with figures 0.2, 0.3 and 0.4:

1. **internal stimuli** (also known as endogenous stimuli) — these are perceptions of the state of the animal's internal environment, such as perceived internal water deficit, perceived body temperature, etc.
2. **external stimuli** (also known as exogenous stimuli) — these are perceptions of the state of the external world, such as perceived food sources, perceived predators, etc.
3. **indeterminate stimuli** — some stimuli which are relevant to the animal are neither obviously internal nor obviously external. For instance, circadian rhythms in many animals are based on an internal clock which gets continually reset by the external day/night cycle [e.g. Marler & Hamilton, 66]. Similarly, in order for the animal not to get lost, it must have access to some measurement of how well it knows where it is. That is to say, the animal must respond to a stimulus which is related to the variance of its estimate of its own position (see section 3.4). This stimulus (the variance of the animal's estimate of its position) can not be directly obtained from the current state of the external environment, and so it is not an obvious external stimulus. However, it is dependent on the external environment (encountering recognised features) rather than the internal environment. Any stimuli, such as the two just described, which do not fall neatly into either of the first two categories are classed here as indeterminate stimuli.

4. **action** — this is the name given to an entity at the lowest level, such as moving in a certain direction, sleeping or freezing (remaining motionless to avoid detection by predators). It is assumed that actions lie at the level of the 'behavioural final common path' [e.g. McFarland & Sibly, 75]. That is to say, the animal is forced to choose only one action to be executed at any moment in time. Different actions are incompatible in terms of their demands on the motor system of the animal, and are therefore mutually exclusive. Actions and *fixed action patterns* (rigid sequences of actions) are shown on the right-hand side of figure 0.2 (e.g. 'looking around', 'trampling' and 'pecking').
5. **action selection (AS)** — this is the process of choosing, at each moment in time, the most appropriate action with regard to all types of stimuli.
6. **action selection problem** — this is the overall problem for an animal of how to select its actions so as to maximise its future expected genetic fitness by surviving and reproducing, and by helping close genetic relatives to do likewise.
7. **action selection mechanism (ASM)** — this is a computational mechanism which can produce a selected action as output when given different stimuli as inputs. The problem of action selection is the problem of *what* action the animal should choose at each moment in time; an action selection mechanism specifies *how* those actions are chosen.
8. **simulated environment (SE)** — this is a computer program which models the action selection problem for an animal and thereby allows the implementation and evaluation of action selection mechanisms. The abbreviation SE will be used to refer to simulated environments in general. 'The SE' will be used to denote the simulated environment proposed in part I of this thesis.
9. **sub-problem** — the overall action selection problem creates other subsidiary problems. For instance, the need to survive creates subsidiary problems of maintaining food, water and temperature levels within certain bounds, avoiding predators, keeping fur/feathers in good condition, etc.

10. **system** — each sub-problem of the overall action selection problem will have a separate system of the action selection mechanism devoted to it. Systems frequently overlap though. For instance, two systems might both make use of the action 'rest'. System-level nodes are shown on the left-hand side of figure 0.2 (e.g. 'nesting', 'escape' and 'preening'). A system comprises not just the system-level node, but also all other nodes 'beneath' it. It is equivalent to a 'sub-tree', not a node in a tree.
11. **sub-system** — this is intermediate in size and abstraction between an action and a system. Some sub-systems in figure 0.2 are 'settling', 'locomotion' and 'trimming'.

The terms system, sub-system and action, as used here to describe entities on the 'solution side', are similar to terms used by Baerends [1976], but rather dissimilar to the terminology proposed in [McFarland & Sibly, 75]. These differences are discussed further in section 5.4.

## Overview

Part I of the thesis describes a simulated environment that was developed to enable the experimental testing of action selection mechanisms. Part II considers mechanisms for action selection and contains several contributions:

1. the simulated environment is used to test and compare several different action selection mechanisms. Shortcomings of various mechanisms are described, and theoretical analyses explaining each inadequacy are presented. Similarly, differences in performance between mechanisms are described and explained.

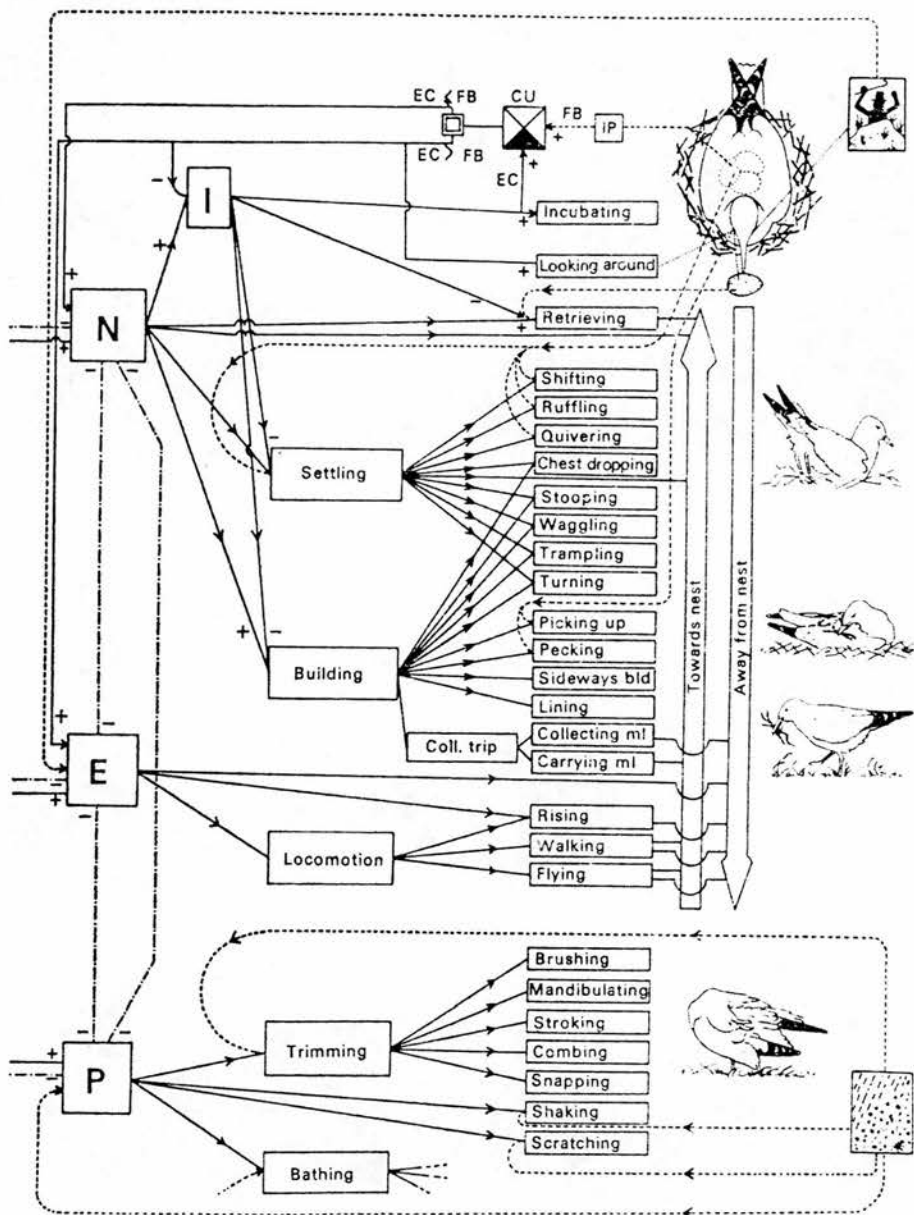


Figure 0.2: Baerends' model to account for incubation behaviour of the herring gull [Baerends, 76]. Actions or fixed action patterns are in the far right column. "Superimposed control systems" of higher order (sub-systems and systems) are to the middle and left. N = incubation system, E = escape system and P = preening system. Taken from [Baerends, 76].

2. a list of requirements for action selection mechanisms is given. Some examples of these requirements are: mechanisms must take account of the hierarchical nature of action selection; *decisions* must only be made at the level of the behavioural final common path; stimuli must be represented as real-valued variables; etc.
3. a novel mechanism for action selection is presented.

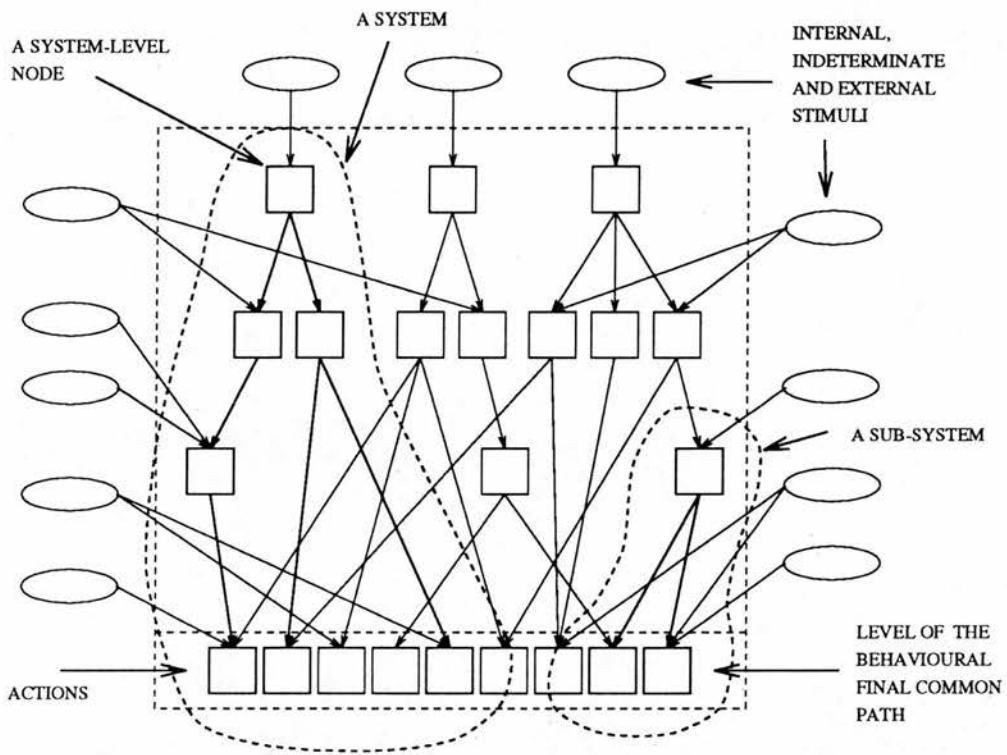


Figure 0.3: An example of the structure of a hierarchical action selection mechanism. The terms system, sub-system and action are illustrated.

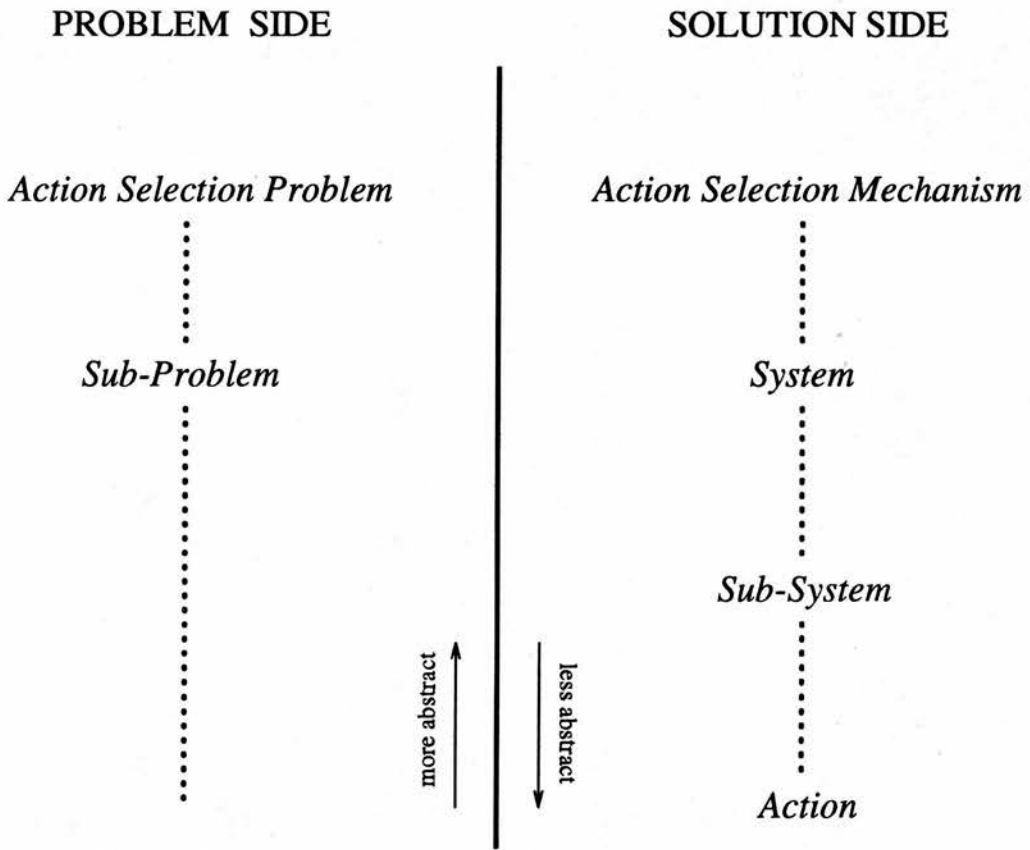


Figure 0.4: Labels for different entities in the description of the problem for the animal, and in the description of the solutions to that problem.

---



## **Part I**

# **A Simulated Environment**

# Chapter 1

## Introduction

Since the time of Charles Darwin it has been accepted that the physical characteristics of animals are the result of evolution due to the processes of natural selection. It is now accepted that the behaviour of animals has also been formed by natural selection.

The part of animals that is responsible for controlling behaviour is the central nervous system. As Albus [81] stated "All brains, even those of the tiniest insects, generate and control behaviour". Although brains can be responsible for other functions (e.g. control of the autonomic system), their predominant function is the generation of appropriate behaviour.

This basic task of animal brains has often been split into three sub-tasks as shown in figure 1.1a. A slightly modified system has been assumed here with the addition of a fourth sub-task, navigation, as shown in figure 1.1b. The four sub-tasks are therefore:

1. Sensing of the environment and interpretation of the sensory signals to provide a high-level description of the internal and external environment at each moment in time (**perception**).

2. Keeping track of where the animal is in its environment, and remembering the positions of important features in that environment (**navigation**).
3. Using the perceptual and navigational inputs to decide which of the animal's repertoire of actions is most appropriate at that moment in time (**action selection**).
4. Transforming the chosen action into a pattern of contractions and relaxations of muscles so as to produce movements and rotations of parts of the body (**motor control**).

Brooks [1986] has argued against a 'horizontal decomposition' of this sort (figure 1.2a), in which there is a sequential, multi-step process in which all sensor information gets fed to one global modelling unit, then on to one global planning unit, and so on. Instead, he argues for a 'vertical decomposition' (figure 1.2b) in which each of a number of task-achieving processes receive and process their own perceptual sensor information, and then calculate their own independent commands for the actuators (motor effectors) or for the lower-level processes.

This SE simulates the perceptual and navigational stimulus inputs to an ASM, and models the effects of the motor commands that are required to execute individual actions, but does not make any assumptions about whether the inputs are computed globally or locally, or whether the motor commands are calculated by global or local entities. That is to say, the SE does not assume either a horizontal or vertical decomposition. There has to be some arbitration between different actions, since otherwise many incompatible actions will be attempted simultaneously and the resultant set of motor commands will be chaotic, but there is no assumption of any other communication between, or global control over, the different systems.

Action selection, along with navigation, is perhaps the least well understood of the four different parts portrayed in figure 1.1b due to the fact that the processes involved are internal and further removed from the outside world than the processes of perception and motor control. Whereas perception can be studied by presenting known stimuli

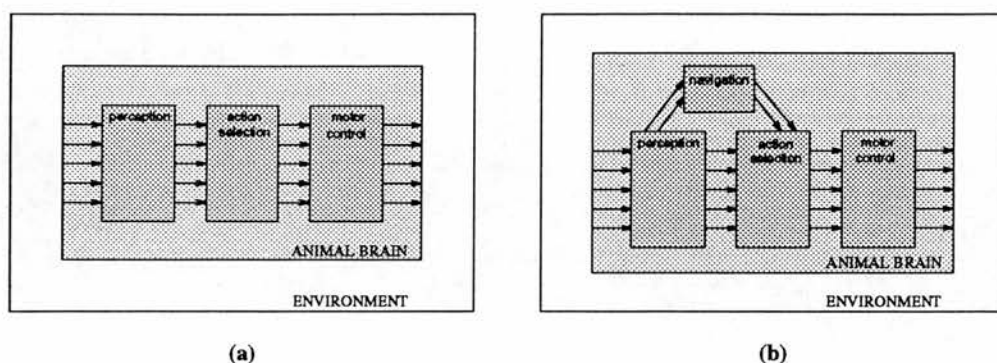
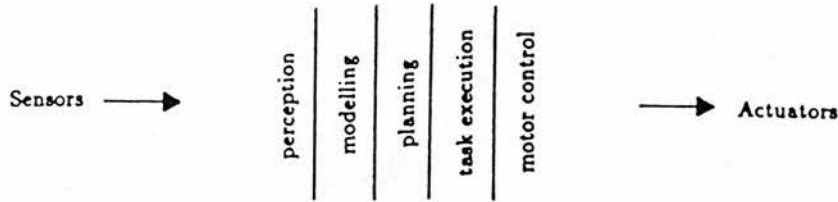


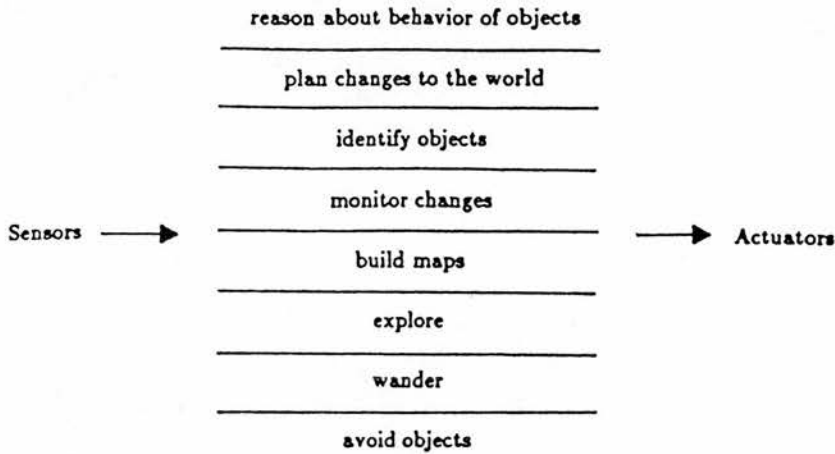
Figure 1.1: Two decompositions of the function of an animal's brain: (a) the more traditional decomposition, and (b) the decomposition used here.

and measuring neural firing rates, and motor control can be studied by exciting neurons and observing motor responses, it is not so easy to apply either of these techniques to the behavioural parts of the brain. Although areas of the brain governing behavioural responses will respond to perceptual stimuli in certain cases, and produce motor responses, the relationships are more complex because the inputs and outputs only interact with the outside world via other interfacing systems [Halliday, 83].

Because of these difficulties in understanding the physiology of AS, theories of how AS takes place in animals have generally existed 'in a vacuum'. Many theories have been proposed, but, because there are no means of validating or invalidating them, little progress has been made. The theories cannot be compared to what occurs in animals because the physiology of the appropriate parts of animals' brains is little understood. The theories have not been able to be implemented as mechanisms whose performance can be observed (to see if they can reproduce the behavioural phenomena produced by animals) because no satisfactory means of doing so has been available.



(a)



(b)

Figure 1.2: Two possible decompositions of a mobile robot control system, according to Brooks. (a) shows a traditional, horizontal decomposition; (b) shows Brooks' preferred vertical decomposition. Taken from [Brooks, 86].

## 1.1 Testing Theories of Action Selection Using Robots

One response to this problem of how to test the validity of ASMs has involved the construction of robots which can 'behave' in (usually highly-constrained) artificial environments [e.g. Peabody, 91 or Rosenblatt & Payton, 89]. A particular ASM can be implemented and allowed to control the behaviour of a robot and the efficacy of the resulting choices of actions can then be observed. While this approach does allow for some testing and comparison between ASMs, there are some limitations to it.

It is now generally accepted that the problems of getting a robot to perceive and

manipulate its world satisfactorily are enormously difficult [Brooks, 87 & 90]. It used to be the case that AI researchers concentrated on the more 'intellectual' aspects of thought, assuming that perception and motor control were trivial and could easily be solved later. Nowadays, however, perception and motor control are recognised as extremely hard problems. Little progress has been made in getting robots to achieve them effectively, at least when compared to the abilities of higher animals.

Because robots are not very good at interpreting complex environments, they need to be placed in environments where there are few features, each of which is easily distinguishable. Because robots are generally rather bad (compared to animals) at coordinating complex motor responses, they will usually possess only a limited repertoire of possible actions, most of which revolve around locomotion. A typical research robot can move effectively around a flat, fairly uncluttered environment but can do little else in terms of manipulating its environment.

Because of these problems with robots, using them to investigate AS in complex AS problems is difficult. While recognising this, it is also worth pointing out that the use of robots has a methodological advantage in that they are grounded in the real world, whereas simulations are not. This is a benefit because it means that people using robots are less likely to make incorrect assumptions about the inputs and outputs of an ASM. Roboticists *know* that perception of the environment is error-prone, that intended actions are not always accomplished as desired, and so on.

In summary, while the use of robots to investigate ASMs is valid, they are at present limited in their usefulness by the simplicity of the AS problems that they can be presented with. This will become less of a drawback in the future as sensor and effector mechanisms become more advanced. At the current time though it is hard to present a robot with a complex AS problem composed of many different sub-problems which vary in many different ways, and for which there is a large repertoire of possible actions to choose from.

## 1.2 The Value of Computer Simulations

The human brain has a limited ability to conceptualise complex dynamic processes such as the interactions between a complicated ASM and the rapidly changing internal and external environment provided by the complex simulated environment described here. It is not possible for people to predict accurately how such processes will change with time. Other examples of systems which are too complex to be fully predictable by introspection alone are the patterns of flow of air over various shapes (e.g. aerofoils) and the behaviour of complex industrial installations such as nuclear power stations under varying conditions. In cases such as these there are a few techniques that can be applied to help with the visualisation or calculation of the behaviour of the system: (i) mathematical analysis – often intractable for more complex systems, (ii) construction and observation of a simplified, smaller-scale physical model – not always feasible or realistic enough, and (iii) design and observation of a computer simulation. Mathematical analysis is not feasible for the interactions between a reasonably complex AS problem and a non-trivial action selection mechanism (see section 5.5). A simplified physical model in this case corresponds to a robot.

Although there are drawbacks to the use of a simulation to investigate action selection (see chapter 5), there are also three important benefits:

1. using an SE allows the 'bypassing' of the sensory perception and navigation processes (the relevant information is calculated by the SE and passed to the ASM), and also of the motor control process (the effects of the animal's actions on the environment are calculated by the SE computer program without concern as to how the actions are achieved). The perception, navigation and motor control systems are all part of the SE. They only have to be programmed once, after which it is possible to just slot in different ASMs without adjusting any of the rest of the programming (see figure 1.3).

2. In order to simulate an ASM, it must be implemented, which requires an explicit and complete specification of the mechanism. The process of implementation brings to the fore any vaguenesses, ambiguities and inconsistencies in the description of the mechanism. Theories of action selection can be described at a fairly high level, and may sound perfectly sensible on paper, but turn out in practice to be rather vague and completely unimplementable as mechanisms.
3. The process of simulating an animal environment, and then simulating a mechanism which selects actions in it, can, if the simulation captures the important characteristics of real animal environments, give some idea of how the mechanism could perform in the real world. Most importantly, gross defects in performance, such as completely inappropriate selection of actions in this case, will show up in the simulation.

With respect to point 3, it will be argued (section 5.3) that defects or differences in the performances of ASMs in the SE are not important in themselves. They are only useful as indicators of phenomena which require further investigation. The defective performance of an ASM is only useful when it can be shown that it would also occur in the real world, and is a product of a fault in the way the mechanism selects actions rather than a product of some quirk of the programming of the SE. Similarly, differences in performance between mechanisms are only useful when the reasons behind them are understood.

### 1.3 Other Work on Simulated Environments

Other simulated environments have been written before, but up to now none with the complexity of that presented here. A brief review of these other environments is now given:

1. **RAM** [Taylor *et al*, 89]: this SE was designed for the purpose of examining population behaviour and dynamics, such as the variation in population sizes of



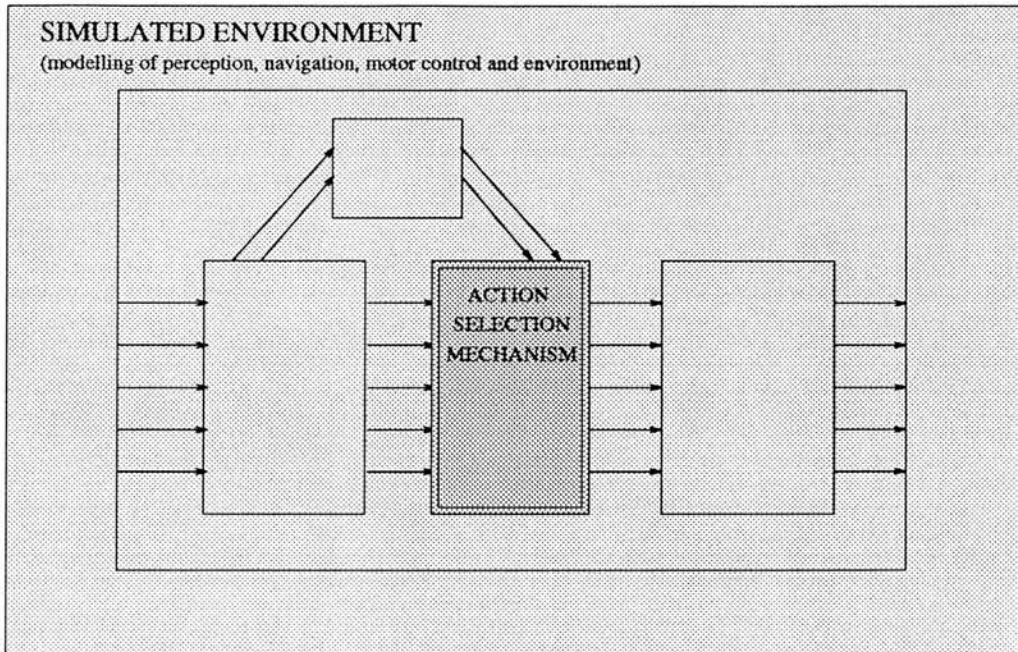


Figure 1.3: Relationship between the simulated environment and the action selection mechanisms. The environment (static and dynamic features, weather, day/night cycle, etc) and perception, navigation and motor control are all calculated in the simulated environment. Different action selection mechanisms can be slotted in to the simulated environment.

predators and prey with different strategies, and the formation of 'leks' (arenas/display grounds where males congregate and defend neighbouring patches of territory on which to display to potential mates) by birds such as the sage grouse (see figure 1.4). RAM was constructed with the intent of examining the population effects of various AS strategies, rather than evaluating the efficacy of an ASM for an individual animal. There are animals in RAM and it is possible to add other features if desired. The squares of the grid can have different attributes such as 'food value', or 'altitude', and there are global environment variables such as 'time of day'. The perception of the animals is local and noise-free. The

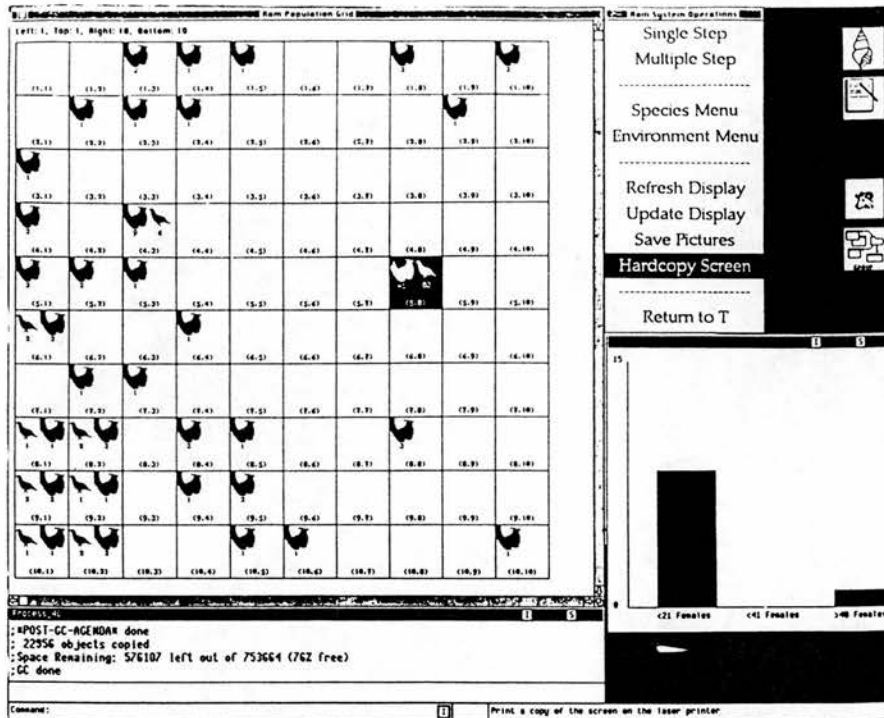


Figure 1.4: RAM. The numbers underneath the male and female birds indicate how many of them are in the square. A 'lek' has formed at (5,8). Taken from [Taylor *et al*, 89].

only possible actions are movements.

2. **WOODS7** [Wilson, 85]: Wilson studied a simple method for associating perceptual situations and actions that could easily be optimised using classifier systems. The WOODS7 environment (see figure 1.5) does not provide a complex and interesting test of ASMs in general. There are only two types of feature (tree and food) in the environment and perception of these is perfect if they lie in neighbouring squares. There are 8 possible actions — a movement in each of 8 directions to any of the neighbouring squares.
3. **Petworld** [Coderre, 89]: this SE (see figure 1.6) is a "system for modeling non-species-specific behavior". It was designed partly with educational uses in mind, and therefore needed to be of limited complexity. The intended function of Petworld is similar to that of the SE in this thesis — the testing of ASMs. There

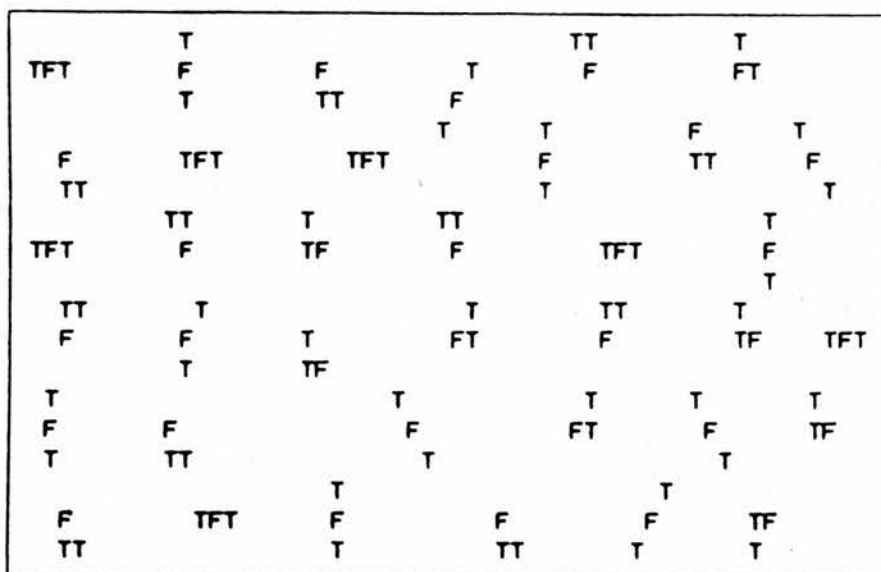


Figure 1.5: Wilson's WOODS7 environment. T stands for tree, F stands for food. Taken from [Wilson, 85].

are two types of feature (apart from the animals themselves): rocks (to be used for building nests) and trees (food). The animals have a limited 'field of view', although it is unstated whether or not perception is perfect within that. There are 13 different actions, 8 of which are movements. There are three internal variables, *hunger*, *fear* and *injury* (damage).

4. **SE to test Maes' mechanism** [Maes, 91a]: this SE (see figure 1.7), written by Jan Torrele, was used for testing Maes' ASM (see chapter 8). The space is not quantised into squares, and so features exist as shapes in the environment rather than being attached to squares. The only types of feature are obstacles, food and water, apart from other animals. Perception is noisy and local. The animals can eat, drink, sleep and fight, as well as being able to move about the environment. The animal has at least two internal variables (food and water levels).

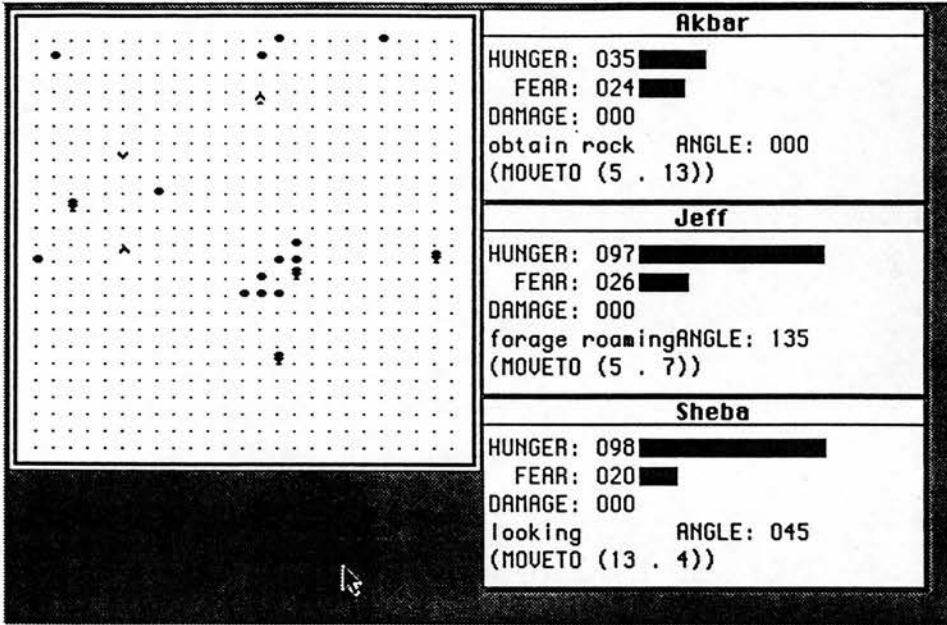


Figure 1.6: Petworld. Individual animals are summarised on the right-hand side. There are three types of feature and three types of internal variable. The 'v' shapes are animals, the ellipses are rocks and the other shapes are trees. Taken from [Coderre, 89].

5. **AL** [Ackley & Littman, 92]: this SE (see figure 1.8) was part of an experiment to examine the interactions between short-term learning (during an animal's lifetime) and long-term learning (over an evolutionary timescale). The evolutionary learning was provided by genetic algorithms which manipulated the initial weights of a neural network. The neural network controlled the behaviour of the simulated animal in AL. Some of the weights in the network could change with experience (short-term learning). The space of AL is quantised into squares. There are 'carnivores' (predators), plants (food), trees (protection from predators) and walls (obstacles) in addition to other animals. Perception is local but perfect, and is only of squares in the four compass directions N, E, S and W. The animal has the two internal variables (health and energy) and four possible actions (movements in the four compass directions).

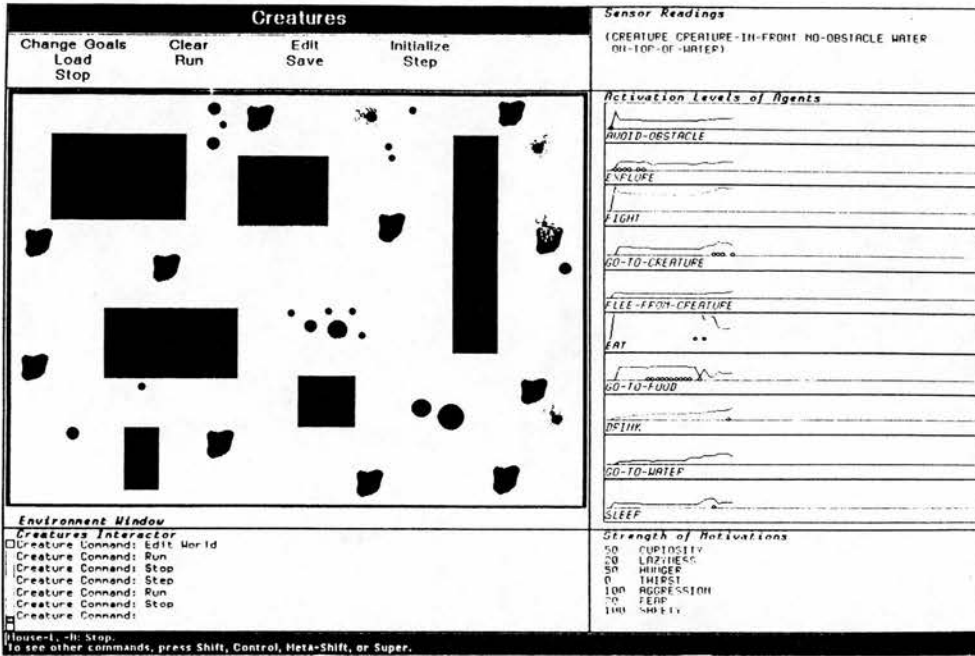


Figure 1.7: SE to test Maes' ASM. The main window on the right shows motivations of the creature over time. The main window on the left shows the environment. Taken from [Maes, 91a].

Some generalisations can be made as to how the five SEs above address the important issues involved in generating a simulated environment. All five of them quantise time into a series of discrete 'ticks' or 'time-steps'. Four of the SEs are grid-based and have features which are attached to squares in the grid. The other one has a continuous representation of space. The first has variables for each square of the grid, whereas in the others there is only the presence or absence of features, which can themselves be variable. All of the SEs possess only a very few different types of feature (a maximum of four in AL).

In all of these SEs except the fourth the animal receives perfect knowledge (has perfect perception) of its local surroundings. The execution of actions is also infallible. If an animal decides to execute an action then there is no question about whether or not it will be able to do so. The animals can typically do little more than move about their

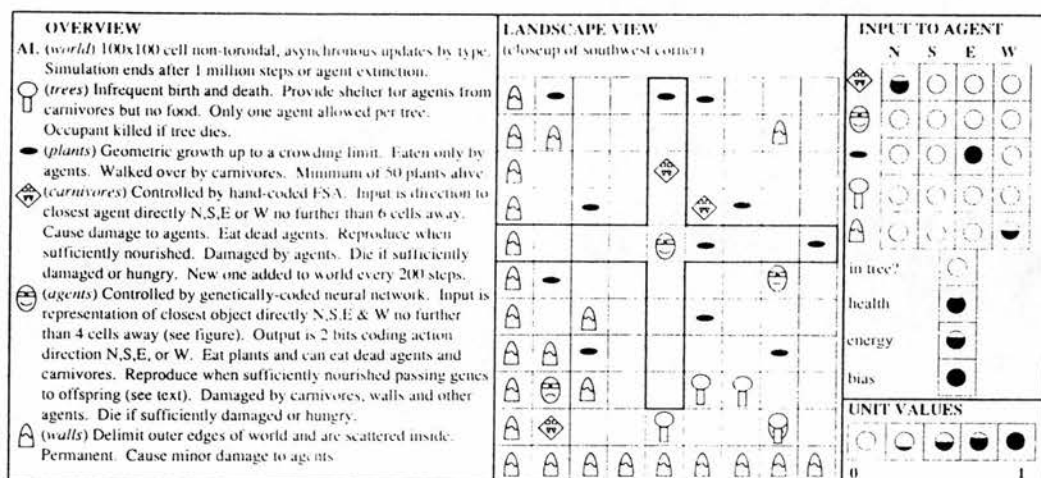


Figure 1.8: AL. The diagram contains a description of the SE, a close-up of part of the SE, and the perception and internal state of the animal. Taken from [Ackley & Littman, 92].

environment, with the maximum number of actions being 13 in Petworld. There are usually two or three internal variables but no more.

There is very little theory that is directly related to SEs. One relevant paper by Wilson [91] outlines a tentative taxonomy for SEs. Some ways in which he suggests SEs can differ are (i) average length of time between the correct action in response to a stimulus, and the associated reward, (ii) reliability of stimuli as an indicator of reward (e.g. will certain stimuli always be followed by rewards if the correct actions are taken, or only sometimes), (iii) regularity of environment characteristics (e.g. do features of the same type vary, or are they always identical), (iv) amount of noise introduced to perception, and (v) amount of noise introduced into reward.

The five SEs described above are all fairly simplistic and lie at the lower end of most of the scales of measurement suggested by Wilson (for example, reward generally



occurs instantaneously, stimulus is a reliable indicator of reward, and there is no noise in either perception or reward signals). This is partly because the SEs haven't been designed to test action selection itself, and also partly because the computing power required for more complex simulations has only recently become widely available. The SE to be presented in this first half of the thesis is more complex than any of those so far outlined and is a more rigorous attempt to capture the nature of the action selection problem for animals. There are more internal variables, more different types of features (with variability between features of the same type) and many more actions. Both perception and motor control are fallible and the model of perception is much more sophisticated than any of those described here. In addition, a fairly sophisticated model of error-prone navigation is included.

The relationship between stimulus and reward, which is stressed in Wilson's taxonomy, is not so pertinent for this SE since it is not designed to test learning; all the mechanisms tested will be 'hard-wired', with no learning component.

## **1.4 Plan of Part I**

Part I of the thesis focuses on the SE and contains five further chapters: Chapter 2 presents the theory and issues taken into consideration when designing the SE. Chapter 3 describes the SE. Chapter 4 explains the testing procedure used to ensure that the programming of the SE was 'bug-free' and shows that the SE presents an interesting challenge to ASMs. Chapter 5 discusses some limitations to the usefulness of the SE and considers the implications of some of the assumptions made during its construction, and finally Chapter 6 contains some concluding remarks and suggestions for further work.

## **Chapter 2**

# **Theory**

This chapter contains some theoretical considerations that were taken into account during the design of the SE.

### **2.1 Prior Assumptions**

In writing any simulation it is usually necessary to make simplifying assumptions, in order to make the simulation computationally tractable. What can make or break the value of the simulation is the validity of these assumptions. An attempt was made to minimise the number of prior assumptions in this work, and to try and restrict those that had to be made to safe and incontestable ones.

The SE in this thesis was designed as a 'testbed' for ASMs. It was important that during the writing of the SE, no preconceptions about the form of the ASMs to be tested were built in. In order to try and achieve this, the writing of the SE was carried out before most of the literature on ASMs was studied. The result is, hopefully, a testbed with no inherent biases towards any particular mechanism or solution.



## 2.2 Complexity and Realism

When deciding how complex and realistic to make the SE, several factors had to be taken into account. A large degree of complexity is desirable, as well as a range of different types of sub-problem (see section 7.2), in order to make the challenge to the ASM harder and more realistic. On the other hand, the dynamics of the SE need to be to some degree comprehensible and transparent so that an observer can grasp what is going on and have some idea of how appropriate different actions are. There was also a limit on the amount of time that could be spent programming the SE, and this also constrained the amount of complexity that was achievable. The SE was made as complex and realistic as these two opposing constraints allowed.

## 2.3 Genetic Fitness

As explained in [Dawkins, 89], an animal is just a vehicle for the propagation of genes. The genetic fitness of an animal is a measure of the number of copies of its genes it manages to pass on to future generations. There are two ways in which an animal can propagate its genes. Firstly, it can reproduce as often as possible so as to bring about the creation of many new individuals which share a large proportion of its genes. This goal implies a subsidiary one of surviving as long as possible so as to be able to procreate many times. Secondly, the animal can help other animals which share many of its genes ('close genetic relatives' such as siblings, parents, offspring) to survive and reproduce as many times as possible.

The function of an ASM is to produce the most appropriate action at each moment in time, where the 'most appropriate' is that which maximises the future expected genetic fitness. *Future* expected genetic fitness is relevant here because it is not always optimal for an animal to choose the action which it is expected will maximise genetic fitness in the short term. For instance, if the animal perceives a predator and a mate simultaneously then its higher priority should probably be to avoid the predator so as to be able to survive and mate in the future. Maximising short term genetic fitness

would lead the animal to try and copulate with the mate, even if that incurred a high risk of death from predation. It is future *expected* genetic fitness because the interactions between an animal and a real environment are not deterministic from the point of view of the animal. The animal cannot calculate the exact effect of each potential course of action on its future genetic fitness. It can only estimate the effects of different courses of action.

How can the ability of an ASM to maximise the future expected genetic fitness be measured in an SE? It is not possible to evaluate the suitability of each decision at each moment in time (unless the suitability of each possible action in each possible state of the SE is known in advance). It is only possible to measure the suitability of decisions indirectly by measuring the final genetic fitness at the end of the animal's life. The final genetic fitness will give an indication of the appropriateness of all the selections of actions made during the animal's life.

In a realistic SE there will be an element of randomness, and it will be necessary to determine the average genetic fitness (from many runs in the SE) in order to obtain an accurate measure of how well the animal maximises its future expected genetic fitness at every moment in time in every run. In the SE described here it is possible for a good ASM to obtain a low genetic fitness on a single occasion because, for example, the particular instance of the SE has most of the water sources distributed a long way from the animal's den. For this reason, an average measure is used here to give an indication of the performance (ability to maximise future expected genetic fitness) of an ASM.

However, there is a problem with the intended calculation of the performance of an ASM in the SE: how is it possible to quantify the contribution an animal makes to the number of copies of their genes its close genetic relatives manage to pass on to future generations? Two possible options were available: (i) ignore the interactions an animal makes with close genetic relatives and thereby make the measure of genetic fitness a much simpler one (just the number of times the animal itself reproduces), or

(ii) include close genetic relatives in the SE and include some rather arbitrary measure of how much the animal helps close genetic relatives to pass on their genes (and also take into account in each case their degree of genetic relatedness).

The first of the two options above was chosen. That is to say, *individual* fitness rather than *inclusive* fitness was measured. One reason for this was so as to make the calculation of the genetic fitness simpler and less arbitrary. If an animal cannot increase its genetic fitness through altruism to close genetic relatives then there is no doubt that the only way in which it can do so is through reproducing, and so the genetic fitness of the animal is proportional to the number of times it mates. The second reason is linked to the need to avoid building in preconceptions to the SE. If close genetic relatives of the animal are to be a part of the SE then the SE will include assumptions about the behaviour of conspecifics (members of the same species) of the animal to be tested, which is very close to including assumptions about the behaviour of the animal itself. The SE is a testbed for action selection mechanisms, and as such it is not desirable for it to include any assumptions about the behaviour (= selection of actions) of conspecifics.

Ignoring social interactions in the SE makes a fairly large simplifying assumption. However, the assumption is necessary in order to obtain a principled calculation of genetic fitness and in order to get around the problem of having to model the behaviour of conspecifics in something being used to test the behaviour of the animal. Ignoring social interactions means that it is assumed that the animal in the SE does not come into contact with close genetic relatives, and spends little time interacting with others of its kind (except when mating). This assumption is not realistic for most animals, but in fact holds fairly well for some. Many animals (e.g. snakes, hedgehogs, tigers) lead lives in which they contact others of their kind only very infrequently. The assumption is more likely to be realistic for male animals since they mate but don't usually have to care for the young. It is worth noting that even when an animal spends a lot of its time in social interactions then it will still require a mechanism for choosing between different actions at each moment in time, even though the repertoire of actions will include some extra alternatives (e.g. mutual grooming, threatening, suckling, protection of

offspring) and the likely benefits of actions will depend on other, less definite factors such as the characters and emotional states of other animals.

## 2.4 Sub-Problems

What the animal has to do in order to increase its genetic fitness has now been defined. It must try and mate as many times as possible during its lifetime. The problem of survival though remains to be defined. What must the animal do in order to stay alive? The definition of genetic fitness has ensured that any ASM will need to have a system for 'reproduction'; the definition of the problem of survival will determine what other systems will be required.

What systems or behaviours are fairly common for animals? McFarland [85] mentions aggression, alarm (signaling the presence of predators), camouflage (of the animals themselves or of their nests), communication, courtship, drinking, feeding of self, feeding of young, foraging, hibernation, incubation, migration, navigation, nesting, predator avoidance, predation, reproduction, suckling, territorial defence and vigilance (i.e. scanning the environment for predators). The main behaviours covered in [Morris, 90] are appeasing, caring for young, cleaning, courting, drinking, escaping (by freezing, fleeing or fighting), food-finding (by foraging, hunting or scavenging), grouping (combining into social groups), mating, nesting, playing, preparing food (e.g. removing shells), sleeping, storing and relocating food, and tool-using. Dewsbury [1978] lists locomotion, ingestion (including feeding, drinking, breathing), thermoregulation, seeking of shelter, avoidance of predators (including concealment, warning of conspecifics, escape and fighting), sleep, body maintenance (i.e. cleaning), elimination (urination and defecation), exploration, play, use of tools, reproduction (including courtship, mating, care of eggs and care of young) and social behaviour (including dominance, territoriality, aggression and social facilitation). Kilmer *et al.*, [69] list the following "modes of vertebrate behaviour": building or locating the nest, defecating, drinking, eating, fighting, fleeing, giving birth, grooming, hunting (for prey or fodder),

mating, mothering the young (including suckling or hatching, retrieval, perineal licking and so on), searching (or exploring), sleeping, urinating, and special innate forms of behaviour such as migrating, hibernating, gnawing, hoarding, etc depending on individual species.

Because it was decided to keep social interactions to a minimum and the animal was assumed to be a male the following sub-problems were ignored: aggression to conspecifics, appeasing, communication (e.g. alarm behaviour, language), dominance, giving birth and care of young (including distracting, feeding of young, incubation, nesting, suckling), grouping, playing, social facilitation and territorial defence. Some others were discarded because they are fairly specialised behaviours that are relevant only to a subset of animals: camouflage, death-feigning, hibernation, hoarding, migration, mobbing, preparing food, startling, storing food and tool-using. This leaves a list of fairly common, mostly non-social sub-problems which are:

**cleaning** (preening, grooming of self, washing, dust-baths, etc).

**obtaining food** (including foraging, eating).

**obtaining water** (including looking for water, drinking).

**temperature regulation** (including resting, finding shelter/shade).

**predator avoidance** (including running away, heading for protection, freezing).

**vigilance** (looking about for predators).

**sleeping at night** (including heading back to den, sleeping).

**reproduction** (including courting, copulating).

Another four sub-problems were considered to be relevant, although they do not feature in any of the four lists above (perhaps because they are not such easily distinguishable entities and do not have their own unique associated actions or fixed action patterns). These are:

**hazard avoidance** (e.g. most animals need to avoid certain areas of the environment such as vertical drops, streams/rivers or dense undergrowth).

**irrelevant animals avoidance** (animals which are not predators but which might injure the animal if it does not keep out of their way).

**not getting lost** (the animal needs to maintain an accurate estimate of where it is so that it can find its den at night and can navigate successfully to remembered places).

**staying close to cover** (so that the animal can escape from predators when they appear).

These four sub-problems are not commonly included in lists of animals' behaviours because they do not have specialised actions or fixed action patterns associated with them — they just affect the direction the animal moves in. Nor do they produce any distinct sequences of appetitive and consummatory actions. Nevertheless, they are sub-problems animals must take into account when deciding the best action to select, and there will need to be systems dedicated to them.

It should be noted that the three systems 'predator avoidance', 'vigilance', and 'staying close to cover' are not collapsed into one anti-predator system because there are different immediate causal factors in each case (the perception of predators, the time since the last scan for predators and the distance from cover). This issue is discussed further in section 5.4.

A further sub-problem, **edge avoidance**, was required to stop the animal from leaving the finite SE.

This choice of sub-problems dictates the features present in the SE, the internal variables of the animal and the factors that impinge on the animal's health (all described in chapter 3). The exact choice of which sub-problems to include is rather arbitrary, but



fortunately the exact choice is not overridingly important. It *is* important though to have a complex SE with a large assortment of sub-problems in order to make the action selection problem interesting. It is also important to have sub-problems with different properties (see section 7.2). This has been achieved with the choice above. For instance, some sub-problems are homeostatic, some are not; some are external stimulus dependent, some are not; some are proscriptive, some are prescriptive; etc.

## 2.5 Modelling Perception, Navigation and Motor Control

In section 1.2 it was mentioned that one of the benefits of using a simulation was that perception, navigation and motor control could be included in the SE, and the ASM could be considered separately. The properties of perception, navigation and motor control can all be modelled at a fairly abstract level with no need to worry about the lower-level mechanics as long as the higher-level characteristics relevant to action selection are realistic. A discussion of the models used in the SE is given here. More details for each case are given in sections 3.3, 3.4 and 3.5 respectively.

1. **perception** – this process transforms a large amount of low-level sensory data (e.g. signals from numerous different smell receptors, signals from retinal rods and cones) into a higher-level description of the environment in terms of features and their positions. Animals use a wide variety of different senses to obtain information about the environment (e.g. sight, sound, taste, smell, touch, echolocation, heat-detection). The main assumption made here is that the animal, using whatever senses, can only sense a fairly restricted area around it (i.e. a local area of the environment). The model used here is of an error-prone process. For every feature in the area of the SE local to the animal, a probability of incorrect perception is calculated. If a randomly distributed number between 0.0 and 1.0 is less than this probability then the perception of the feature is distorted and the animal will either perceive nothing there or mistakenly perceive a different type of feature. In some cases the animal will also perceive features where none exist.

The probability of incorrect perception is affected by intervening vegetation, distance from the animal, the selected action of the animal, and the time of day.

The animal can also sense the values of its internal variables (e.g. food and water deficits and body temperature), again with a degree of noise built into the perception. Finally, the animal has access to the values of various indeterminate stimuli.

2. **navigation** – the term navigation is used here to cover the creation and usage of a ‘map’ (a collection of memories of the positions and attributes of various features), as well as the estimation by the animal of its current position, and the calculation of how to move towards a remembered feature. The animal forms error-prone memories of features based on its perception of them and on the animal’s estimate of its own position when it encounters them. The variance of the animal’s estimate of its position increases each time it moves, but can decrease when the animal encounters recognised features. The memories will become more accurate as the animal visits the features more often.

It was stated in the Prologue that none of the mechanisms incorporate learning. Navigation is not an exception to this rule. Although the navigational inputs to an ASM will change through the animal’s lifetime as it becomes more acquainted with the layout of its environment, this does not mean that any learning takes place within any of the ASMs.

There is a limit to how many features can be remembered at any one time. The strength of a memory is dependent on the utility of the feature, the number of times it has been visited and how long since it was last visited. Lower strength memories can be removed from the map (be ‘forgotten’) as newer, stronger ones are added. The animal can use its memories of features to head towards a remembered feature, though the likelihood of finding it depends on how well it has remembered the feature’s position and on how well it knows its own position, as well as on the existence of hazards between the animal and the feature.



In summary, the map consists of error-prone memories which can be forgotten over time and which can not always be used successfully to find the feature they represent.

3. **motor control** – this process models the transformation from a chosen action into a set of lower level motor commands that bring about movements of the animal's body. In the model used here the process is once again error-prone. There is a fixed probability of an action not being successfully executed. This probability only increases if the animal is incapacitated (has very bad health). The low-level details of motor control are not simulated, only the end effects of each action (e.g. an increase in the level of the animal's internal water and a decrease in level of a water source after the animal drinks there, a change in the position of the animal and an increase in the animal's internal temperature when it moves fast).

## Chapter 3

# Description

Figures 3.1 and 3.2 show two randomly generated instances of the SE (figure 3.3 explains the meanings of the different symbols). The SE consists of a  $25 \times 25$  grid of *squares* together with many different *features* which can occupy the different squares. Random numbers of features are distributed randomly around the SE, so that any two instances of the SE are likely to be very different. The SE has a fixed edge, for moving over which the animal incurs a health penalty. The animal is able to perceive the edge of the SE so as to be able to avoid it. The SE does *not* wrap around (e.g. the animal does not appear at the right-hand edge after it leaves the left-hand one). The animal's den is always placed somewhere near to the centre of the SE.

Time in the SE is split up into *timesteps*. Each day in the SE consists of 500 timesteps. Each day is also sub-divided into 6 *parts*, which affect how well the animal is able to perceive the local area of the SE. Night occupies the first  $\frac{2}{12}$  and the last  $\frac{2}{12}$  of the day ( $\frac{1}{3}$  in all). Sunrise occupies the first  $\frac{1}{12}$  after night ends, morning the next  $\frac{2}{12}$ , midday the subsequent  $\frac{2}{12}$ , afternoon the next  $\frac{2}{12}$ , and sunset the final  $\frac{1}{12}$  before night.

Most variables in the SE (e.g. attributes of features, animal health, animal internal variables) are scaled to lie in the range 0.0 – 1.0. Unless otherwise stated, a value of 0.0 will correspond to the minimum, and a value of 1.0 will be the maximum.

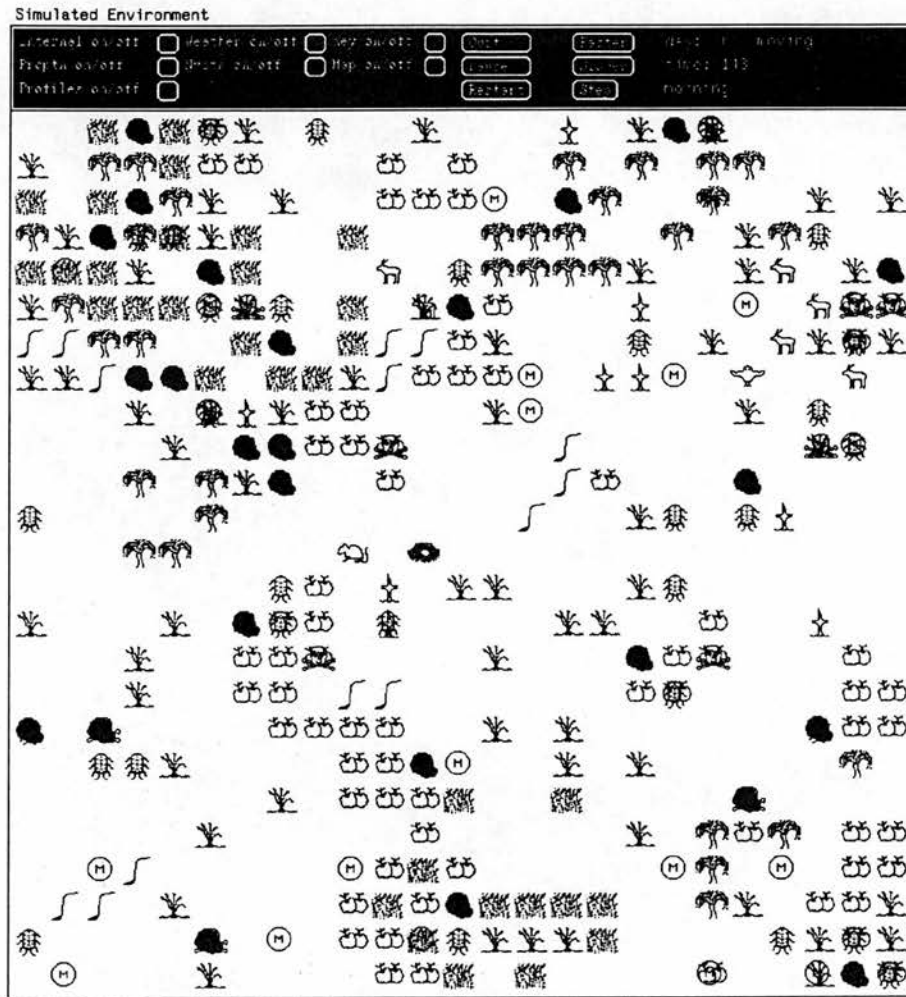


Figure 3.1: A randomly generated instance of the simulated environment showing the positions of features, the position of the animal, the current action of the animal, the time, the day and the current part of the day.

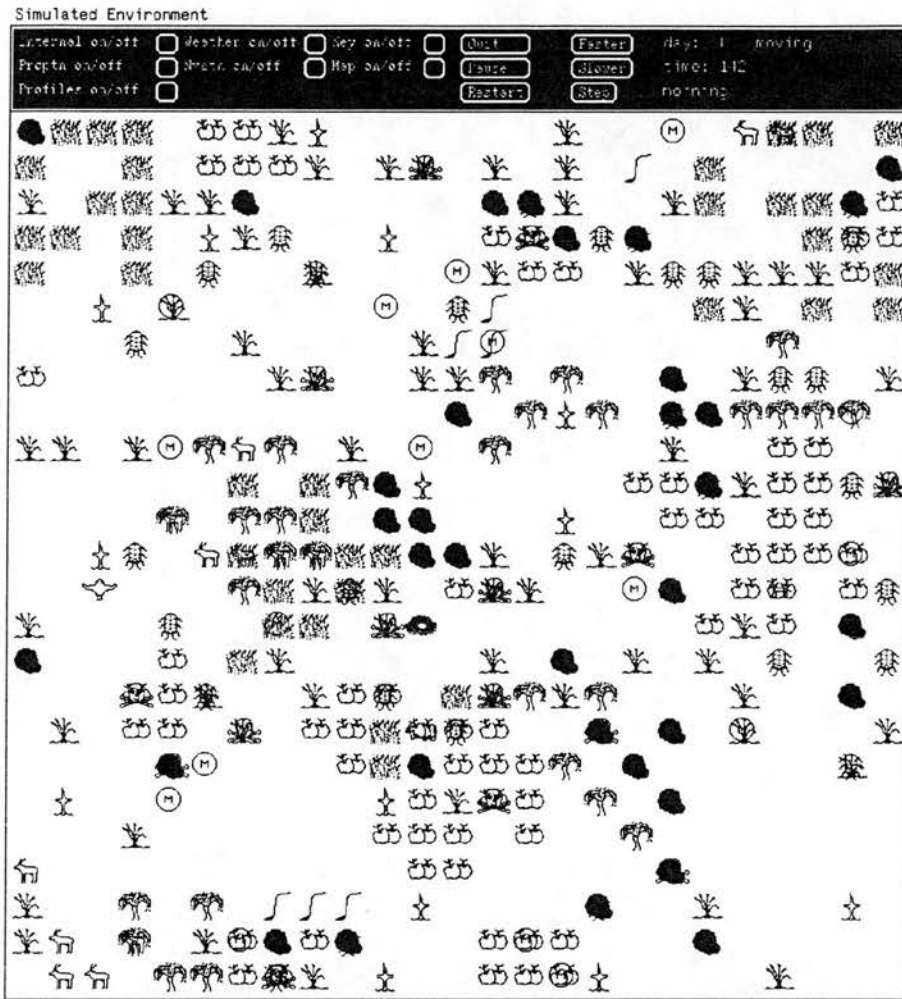


Figure 3.2: A different randomly generated instance of the simulated environment.

The rest of this chapter describes how the different aspects of the SE discussed in the last chapter (genetic fitness, survival, different sub-problems, perception, navigation and motor control) have been implemented in the SE. While a lot of the descriptions here are fairly low-level, in most cases the precise forms of the equations and the exact implementational details are not given, due to the amount of space this would take. An exception to this is the section on navigation. Because it is probably the most contentious of the three processes, and in order to give an example of how one of the

### Key to Feature Types

Lighter shades imply greater value to the animal
















-  - water source
-  - toxic food or water source
-  - cover
-  - shade
-  - dangerous place
-  - landmark
-  - cereal type food
-  - fruit type food
-  - den
-  - irrelevant animal (just needs avoiding)
-  - mate
-  - predator (type 1)
-  - predator (type 2)
-  - prey
-  - animal whose behaviour is being modelled

Figure 3.3: Key to features in the simulated environment.

processes is implemented, a fuller description of navigation is given, including most of the low-level equations used to implement it. The code for this simulated environment can be obtained from the author upon request, in order to repeat the experiments described in this thesis.

While reading this chapter it should be taken into account that although some of the assumptions made here will probably seem rather dubious and arbitrary, the low-level implementational details of the SE are not vitally important as long as the overall high-level 'flavour' of the SE is fairly realistic and presents a challenging test for an ASM. While wanting to keep the SE as realistic as possible in a general sort of way, the SE will only ever be useful as an exploratory tool to give indications of where there are shortcomings with mechanisms. It can never be claimed that it is completely realistic and so it can never be used to prove the existence of shortcomings by itself. Therefore it is not essential that each small, low-level detail be correct.

### 3.1 Animal's Health

In chapter 2 it was argued that the genetic fitness of a solitary, non-social animal is equal to the number of times it reproduces. This leads to two requirements on the part of the animal: (i) reproduction, and (ii) staying alive in order to reproduce. In the SE the survival problem is mediated through the use of a variable *animal\_health* which can vary in the range 0.0 – 1.0. It starts at 1.0 (animal in perfect health) and the animal will die when it falls to 0.0. Different facets are added to the survival problem in the SE by allowing different factors to impinge upon the value of the variable *animal\_health*.

The animal is given a maximum possible lifespan of 10 days = 5000 time steps. This is achieved by imposing an age-dependent upper limit on the value of *animal\_health*. This upper limit stays at 1.0 until the animal has reached  $\frac{1}{2}$  of its maximum lifespan (= 5 days), after which the upper limit is decreased linearly with age until it falls to 0.0 at the maximum lifespan. In practice, even a near-optimal animal will never be able to

live quite as long as the maximum lifespan because other factors will always decrease *animal\_health* some way beneath the upper limit.

As well as 'old age' affecting the upper limit on health, the animal can also suffer *permanent injury* (e.g. from attacks by predators or from hazards). The size of this variable permanently reduces the upper limit of health some amount below the value determined by old age, whereas normal decrements to health can be recovered from over time. If the animal's health is depressed due to normal, recoverable, decrements then it is allowed to recover up to 0.005 (= 0.5%) every timestep, dependent on its currently selected action. Any action other than sleeping will reduce the amount it is able to recover in that timestep. The value of *animal\_health* is therefore determined by whichever of the two following calculations produces the smaller result at each timestep

$$animal\_health_t = 1.0 - O - I - S$$

or

$$animal\_health_t = animal\_health_{t-1} + R$$

where  $O$  = decrement due to old age,  $I$  = non-recoverable decrement due to permanent injury,  $S$  = recoverable decrement from sub-systems (e.g. from shortage of food, lack of cleanliness, recent attack by a predator or whatever) and  $R$  = recovery in health since previous timestep. As an example, consider an animal for which (i) age = 6.5 days, (ii) *permanent injury* = 0.13, (iii) recoverable decrements total 0.21, (iv) previous health = 0.355, and (v) current action = CLEAN (see below). In this case the new value of *animal\_health* is calculated as the lesser of  $(1.0 - ((0.65 - 0.50)/0.50) - 0.13 - 0.21) = 0.36$  and  $(0.355 + (0.7 \times 0.005)) = 0.3585$ , i.e. is calculated to be 0.3585.

If the animal's health is greater than 0.1 then there is a 1% chance of executing each action incorrectly. When the animal's health falls below 0.1 then the chance of it failing to carry out each action correctly increases, and it becomes unable to carry out certain strenuous actions such as mating (it becomes partially incapacitated).

## 3.2 Implementation of Sub-Problems

In section 2.4 a list of relevant sub-problems that should be included in the SE was developed. The way in which they have been implemented is now explained. Any environmental features necessary to model the sub-problems are introduced, together with any relevant indeterminate or internal variables, and their dynamics. The relationships between the sub-problems and *animal health* are described, and any actions that need to be incorporated into the animal's repertoire are presented.

### 3.2.1 Cleaning

This sub-problem models the need of animals in the wild to maintain their feathers, fur or skin in a clean and parasite-free state. Lack of preening or cleaning or grooming can lead to difficulties caused by less effective insulation, infection of wounds, infestation with parasites, etc.

The animal is provided with a low-level action **CLEAN**, and an internal variable *cleanliness* which can vary in the range 0.0 (maximally dirty/dishevelled/parasite-ridden) to 1.0 (maximally clean). At every timestep *cleanliness* is decreased by an average of 0.001. On every occasion that the animal selects the action CLEAN then the difference between the current value of *cleanliness* and 1.0 is reduced by a factor of 0.15 (i.e.  $new\ cleanliness = old\ cleanliness + ((1.0 - old\ cleanliness) \times 0.15)$ ). So if *cleanliness* = 0.2, then successive cleaning actions will increase it to 0.32, 0.422, 0.509, 0.583, etc. The animal's health is reduced by one half of the difference between 1.0 and *cleanliness* (so, for example, a *cleanliness* value of 0.75 means a health decrement of 0.125). Therefore, the maximum possible effect on the animal's health due to lack of cleaning is 0.5.



### 3.2.2 Obtaining Food

Instead of assuming one unitary food variable, three different internal food variables – *fat*, *carbohydrate* and *protein* – are assumed. In real life the need for food is likely to be even more ‘multi-dimensional’, with needs for different minerals, vitamins, etc. However, for simplicity, food is considered in this case to consist of only the three components mentioned above.

Each of the three variables can lie in the range 0.0 (death through lack) to 1.0 (death through surfeit). In practice the values of those food variables are not allowed to reach 1.0, but rather when any of them exceed a value of 0.75, then vomiting is assumed to occur, i.e. the value returns to 0.75 and the animal incurs a recoverable health penalty. When the value of any of the food variables falls below 0.25 then the animal’s health is decreased by  $((0.25 - \text{value})/0.25)^2$  and so the decrement to health increases in proportion to the square of the deficit in the value of the food variable. The value of each of the three variables decreases slowly over time when the animal is not eating. At each timestep the amount of decrease is dependent on the action of the animal, with actions such as moving fast causing higher decrease.

Three separate features are included in the SE for the purpose of modelling this sub-problem: **cereal type food**, **fruit type food** and **prey**. Each of these three types of food differ in their respective contents of fat, carbohydrate and protein, and so different types of food will be more or less useful at different levels of the animal’s *fat*, *carbohydrate* and *protein* internal variables. There are also other differences between the three types of food. Fruit type food and cereal type food instances are both static, with fruit type food instances usually having higher values (i.e. there is more to eat there). The SE has a primitive model of weather, and the values of the cereal type food instances vary in relation to the amount of rain that has occurred in the SE over the last few days. The fruit type food is not related to weather, but is cyclic in nature. Each instance has a ripening-decaying cycle so that the animal will not always find food at a good value fruit food source. Fruit food sources are usually of a higher value than cereal

type food sources (averaged over fat, carbohydrate and protein), and are slightly more numerous. Fruit food sources are more likely to occur near water sources, whereas cereal food sources are positioned completely randomly. Prey type food instances appear at random times on the boundary of the SE and move in an unpredictable fashion across it. If intercepted by the animal then they can be eaten to provide food. They provide a large, instantaneous amount of food.

When the animal eats from fruit food or cereal food sources then their values are decremented by the amount the animal eats. They then 'recuperate' at rates of 0.3 per day and 0.2 per day respectively until they return to their normal values. Instances of prey die when the animal catches and eats them.

Instances of fruit food (10% chance) and cereal food (5% chance) can be toxic. That is, they can induce a decrement in health which will take effect some time after the animal has eaten. The damage to the animal's health will be randomly in the range 0.0 – 0.15 (cereal type food) and 0.0 – 0.40 (fruit type food) and can occur up to several hundreds of timesteps later. The risk of dying from toxicity is not so high as to make the likely penalty of eating greater than the likely reward. The location of food sources can be remembered, and ones that are probably toxic can in future be avoided (see section 3.4).

Three low-level actions are needed in the animal's repertoire: **EAT\_CF**, **EAT\_FF** and **POUNCE** (on prey).

### 3.2.3 Obtaining Water

Unlike food, the need for water is assumed to be uni-dimensional. There is one internal variable, *water*, which can lie in the range 0.0 (death through lack) to 1.0 (death through surfeit). As with food variables, values greater than 0.75 lead to vomiting and values less than 0.25 lead to a decrement in health of  $((0.25 - value)/0.25)^2$ . Again, the amount

that the value of *water* decreases each timestep depends on how strenuous an action has been selected.

An additional feature **water source** is required for the animal to be able to drink from. The value of each water source is dependent on an inherent capacity (set randomly when the water source is initialised) and also on the amount of rain that has occurred in the SE in the last few days.

When an animal drinks from a water source then the water source's value is decreased by however much the animal drank, after which it can recuperate at the rate of 0.03 every 10 timesteps, until it returns to what its value would have been otherwise. Each water source has a 5% chance of being toxic, in which case it will cause a decrement in the animal's health of 0.0 – 0.2 some time later. Water sources can be remembered in the same way as food sources (see section 3.4).

An extra action of **DRINK** is added to the animal's repertoire.

### 3.2.4 Temperature Regulation

The animal is given an internal variable *temperature*, which can vary in the range 0.0 (death due to cold) to 1.0 (death due to heat). If *temperature* is less than 0.25 then the animal's health is decremented by  $((0.25 - \textit{temperature})/0.25)^2$ . If *temperature* is greater than 0.75 then the animal's health is decremented by  $((\textit{temperature} - 0.75)/0.25)^2$ . The value of the internal variable *temperature* is dependent on (i) the external temperature (derived from the primitive weather model and the current part of the day), (ii) the location of the animal (certain features will ameliorate the effects of external temperature), and (iii) the action of the animal (the more strenuous the action, the higher the animal's internal temperature).

Another feature **shade** is used in the SE. This has the effect of protecting the animal against severe external temperatures (either too hot or too cold). The features fruit type

food, cover and den all reduce the effects of external temperature as well. A further action **REST** is made available to the animal. This action leads to a lower than average body temperature. Other actions such as **MATE** and **MOVE\_FAST** lead to higher than average body temperature.

### 3.2.5 Predator Avoidance

Two types of predator features with slightly different characteristics are included in the SE. **Predator1s** move semi-randomly about the SE, and represent ground-based predators. Their movement is determined by balancing four factors: (i) a tendency to move in an initially chosen random direction across the SE, (ii) an attraction towards water sources, (iii) a random component, and (iv) an overriding tendency to chase the animal if it is caught sight of, and if the predator1 is hungry (90% chance). A predator1 can catch sight of the animal from up to 2.5 squares away, whichever part of the day it is. The chance of a predator1 spotting the animal is affected by whether the animal is in vegetation, how conspicuous an action it is executing, and how far away from the predator it is. The predator1 will continue to chase the animal until it catches and kills it or until it gives up. It gives up if it hasn't caught the animal after chasing it for 6–13 timesteps. A predator1 can move between 1 and 3 squares per timestep.

If a predator1 catches the animal (gets in the same square and the animal doesn't escape) then the animal's health will be decremented by a random number between 0.0 and 0.3 (65% chance) or between 0.3 and 1.0 (35% chance). The animal also has a 25% chance of suffering 0.0 – 0.5 permanent injury. In short, if the animal is caught by a predator1 then it has a good chance of being killed outright, whatever its current state of health.

**Predator2s** are similar to a predator1s, but represent avian rather than ground-based predators (e.g. owls or hawks). There are several differences: (i) there are fewer of them during the day and more of them during the night, (ii) they are not attracted

towards water sources, (iii) they move more rapidly (1–4 squares per timestep), (iv) they can catch sight of the animal from up to 3.0 squares away, (v) they give up chasing after 4-8 timesteps, and (vi) they inflict even more damage, on average, to the animal's health if they catch it.

Apart from these two dynamic features, a third feature **cover** is introduced because of this sub-problem. This feature represents thick vegetation and provides protection for the animal against predators. As well as the animal having a better chance of escaping from a predator when in cover, the predator is less likely to spot the animal there. Instances of cover occur in clusters and are more likely to be positioned in the vicinity of water than otherwise. Similarly benefits are given by other vegetation-type features (fruit type food and shade), but to a lesser degree. The animal is completely safe from predators whilst in its den.

Two new types of animal action are added for this sub-problem: (i) **MOVE\_FAST** in a direction enables the animal to try and run away from a predator, and (ii) **FREEZE** (become motionless) makes the animal inconspicuous and, especially when the animal is in vegetation, less likely to be spotted or kept sight of by the predator. The animal's choice of action and the presence or absence of vegetation combine to affect the likelihood of the animal being detected by a predator. So, for instance, the probability of the animal being perceived by a predator is on average 16 times less when the animal is motionless (has chosen the action **FREEZE**) in cover, than when the animal is mating in the open.

When modelling predator-animal interactions, care was taken that predators could not always catch the animal, and also that the animal was not always able to escape predators. The following ordering of events occurs during each timestep:

1. the animal's perception is calculated (including that of any nearby predators).

2. an action is selected by the animal (e.g. moving away from a nearby predator towards cover), and that action is then executed (i.e. any effects on the animal's internal state, or on the features in the simulated environment, are calculated).
3. the perception and movement of predators is calculated (e.g. a predator may catch sight of the animal and move towards it).
4. if any predators have reached the same square as the animal then a check is made to see if the animal is able to escape (dependent on amount of vegetation in square and the action of the animal).
5. if the animal does not escape then it is assumed to have been successfully attacked and its health is decremented by an appropriate, random amount. At this time the animal may also incur permanent injury.

This sequence of events means that the animal can usually escape predators if it perceives them at a distance and takes appropriate action (e.g. moving towards vegetation and freezing there), but that on occasion the animal will be attacked before it has perceived the predator, or will be caught out in the open and will be attacked before it can reach cover.

### 3.2.6 Vigilance

The animal needs to scan the environment for predators every so often, especially when it has perceived one or more predators recently. Two new actions **LOOK\_AROUND** and **LOOK** (in a particular direction) allow the animal to perceive its environment to a greater distance and with improved accuracy. Two new indeterminate stimuli (see Prologue) *time since last scan* and *time since predator last perceived* are used for this sub-problem.



### 3.2.7 Edge Avoidance

When the animal is near the edge of the SE then it will start to try and perceive areas that are outside of it. In this case it perceives a normal type of square, containing only the feature **outside edge**. If the animal tries to move into a square with this feature in, then it is not allowed to and its health is decremented by 0.01.

### 3.2.8 Hazard Avoidance

A new feature of **dangerous place** is used to represent cliffs or streams or bogs, i.e. areas which it is dangerous for the animal to visit. They only have an effect on the animal if it moves into a square containing one of them, in which case the animal has a 40% chance of suffering 0.0 – 0.6 damage to its health, and a 10% chance of suffering 0.6 – 1.0 damage. There is also a 30% chance of 0.0 – 0.5 permanent injury, a 5% chance of 0.5 – 1.0 injury, and a 15% chance of 1.0 injury. In short, there is a very high chance of death or at least very serious injury if the animal enters a square with a dangerous place feature in it.

### 3.2.9 Irrelevant Animal Avoidance

An irrelevant animal is one which does not actively chase the animal but which might injure the animal if the animal doesn't get out of its way. It can be thought of as similar to a moving hazard, needing only to be avoided. If the animal is in the same square as an irrelevant animal and does not manage to escape (the probability of which is dependent on the action of the animal and on the thickness of vegetation) then it will incur 0.0 – 0.4 damage to its health and 0.0 – 0.4 permanent injury.

### 3.2.10 Sleeping at Night

As described in the previous section, predators are able to perceive their surroundings normally at night, whereas the animal's perception becomes progressively worse as night approaches, until the animal can perceive nothing at all outside of its own square. Because of this severely reduced perception at night the animal is also liable to be injured or killed by encounters with dangerous places or irrelevant animals if it moves about, because it cannot see them so as to avoid them. It is therefore highly advantageous for the animal to spend nights in a protected place and engage in actions which do not deplete its resources too rapidly. The feature **den** represents the animal's lair or burrow. The animal is completely safe from predation there. The animal is also protected from extremes of temperature in its den but has limited perception from inside it. There is only one den, near the centre of the SE, and it will be advantageous for the animal to return there every night.

Another action, **SLEEP**, is introduced. This brings about the lowest reduction of food and water per timestep. Two new indeterminate stimuli, *proximity of night* and *distance from den* are included.

### 3.2.11 Staying Close To Cover

It is advantageous for the animal to stay in fairly close proximity to cover (or rather, any sort of protection from predators — fruit type food, shade, the animal's den and cover all offer varying degrees of protection from perception or attack by predators) so that it can head there if a predator is perceived. A new indeterminate stimulus of *distance from cover* is required. There are no new internal variables, features or actions for this sub-problem.



### 3.2.12 Not Getting Lost

This sub-problem depends on another indeterminate stimulus, which is the *variance* of the animal's estimate of its current position. The variance is a measure of how large an error there is likely to be between where the animal is and where it thinks it is. The animal's variance is always set to zero when it is in its den, but then increases as the animal moves further and further away from its den and away from well-known areas (see section 3.4). It is indirectly important for the animal to keep its variance low so that it can find its way to remembered features and back to the den at night. A feature that is added to the SE for this sub-problem is **landmark**, which is an easily recognisable feature which will help the animal to know where it is and to map its environment successfully.

### 3.2.13 Reproduction

This sub-problem is the only one that does not concern survival, as well as the only one involving conspecifics, although of the opposite sex. The feature **mate** moves about the SE and has a 50% chance of being receptive (i.e. 'in heat') at any time (it is assumed to be a female and the animal to be a male). Two new actions **COURT** and **MATE** are added to the animal's repertoire. If the animal enters the same square as a mate and performs the action **COURT** then if the mate is receptive then it will respond to the courting to show that it is ready to mate. If the animal then performs the further action **MATE** then reproduction will be assumed to have occurred and the animal's genetic fitness will increase by 1.0.

If the animal tries to mate with a mate that is not receptive or not courted then the mate will attack the animal. If the animal does not manage to escape then it will suffer 0.0 – 0.25 decrement to health and have a 5% chance of 0.0 – 0.2 permanent injury.

The implementation of all 13 of the different sub-problems has now been described. It can be seen that the whole problem for an ASM in the SE is a complex one. There are 13

different sub-problems involving 14 different features (cereal type food, fruit type food, prey, water, shade, predator1, predator2, cover, den, outside edge, dangerous place, irrelevant animal, landmark, and mate), 6 different internal variables (*cleanliness, fat, carbohydrate, protein, water and temperature*), 6 different indeterminate stimuli (*time since last scan, time since predator perceived, proximity of night, distance from den, distance from cover and variance*) and 35 different actions (8 MOVE actions, 8 MOVE\_FAST actions, 8 LOOK actions, CLEAN, EAT\_CF, EAT\_FF, POUNCE, DRINK, REST, FREEZE, LOOK\_AROUND, SLEEP, COURT and MATE).

### 3.3 Perception

A high-level description of perception was given in section 2.5, the implementation of which will now be explained.

When calculating the animal's perception of the squares around it, 6 factors are taken into account:

1. how far away from the animal the 'target' square is. The maximum distance at which the animal can ever correctly perceive the contents of a square is 3 squares away.
2. the time of day. At night the animal's perception is nil, at sunrise and sunset the effectiveness is reduced by one half.
3. the animal's action. If the animal is purposefully looking around, or purposefully looking directly at the target, then it is more likely to perceive correctly.
4. the presence or absence of vegetation in the animal's square.
5. the presence or absence of vegetation in any of the squares between the animal and the target.



6. the presence or absence of vegetation in the target square (this will not affect perception of the vegetation features themselves, but will affect perception of anything else that is in the same square).

These 6 factors are all taken into account to give a figure in the range 0.0 (0% chance of correct perception) to 1.0 (100% chance of correct perception) that represents the probability that the animal will correctly perceive the given feature. So, for example, the square shown in figure 3.4 might have a probability of being correctly recognised of  $((1.0 - 0.13 - 0.29 - 0.06) \times 0.50) \times 1.0 =$  a total of  $0.26 = 26\%$ , where 0.13 is for the vegetation in the animal's square, 0.29 for blocking vegetation in square X, 0.06 for blocking vegetation in square Y, 0.5 because of the distance of the square, and 1.0 because the animal is purposefully looking in that particular direction.

Once the probability of correctly perceiving the square has been calculated then a random number in the range 0.0 – 1.0 is generated and compared to the probability. If the random number is smaller then the square is correctly perceived (the ASM will receive a faithful representation of what is actually in the square). If it is larger then the ASM receives a corrupted version of what is in the square (i.e. it may not perceive some features which are actually there, and/or it may perceive some features which are not actually there). Incorrect perception will frequently have no effect on the action that the animal would select (e.g. if the animal fails to perceive a fairly sparse instance of cover that is several squares away), but will on occasion have serious consequences (e.g. if the animal mistakenly perceives a predator in an adjacent square and interrupts its approach towards a food source in order to run away from the predator). If the feature has associated perceivable properties (e.g. the 'value' of a water source) then these are corrupted in proportion to the probability of *incorrect* perception. The perception of internal variables is calculated by taking the actual values and adding gaussian random numbers with mean 0.0 and standard deviation 0.002.

Using the techniques just described, the perception of the animal is made to be local and error-prone.

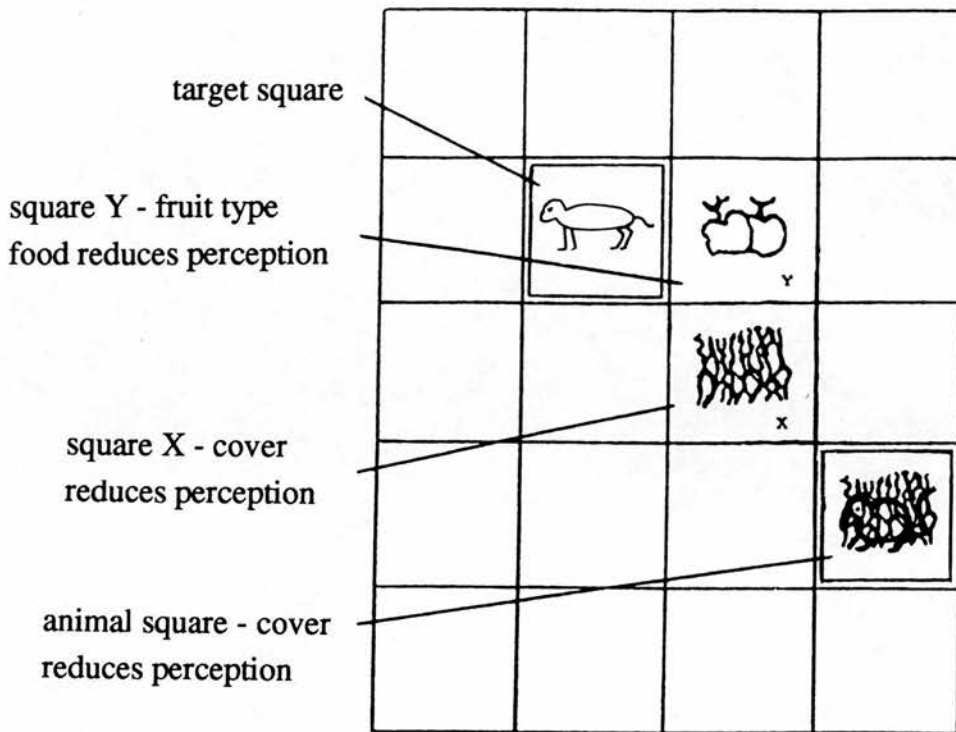


Figure 3.4: Perception example: the chance of the animal correctly perceiving the predator in the target square is reduced by the distance of the square from the animal, vegetation in the animal's square and vegetation in the intervening square *X* and the partially intervening square *Y*.

### 3.4 Navigation

The second of the three parts of an animal brain that need to be modelled in the SE is navigation. This is described in more detail than either perception or motor control. A high level description of navigation was given in section 2.4. A description of how a limited-capacity, error-prone navigation system with realistic properties of remembering and forgetting has been implemented is given below. It must be stressed that the *details* of how navigation is implemented (along with other aspects) do not need to be completely realistic in every respect as long as the gross effect is a fairly realistic model of the process. Most of this process had to be designed from scratch, although some of the equations producing minimum-variance position estimates are taken from elsewhere.

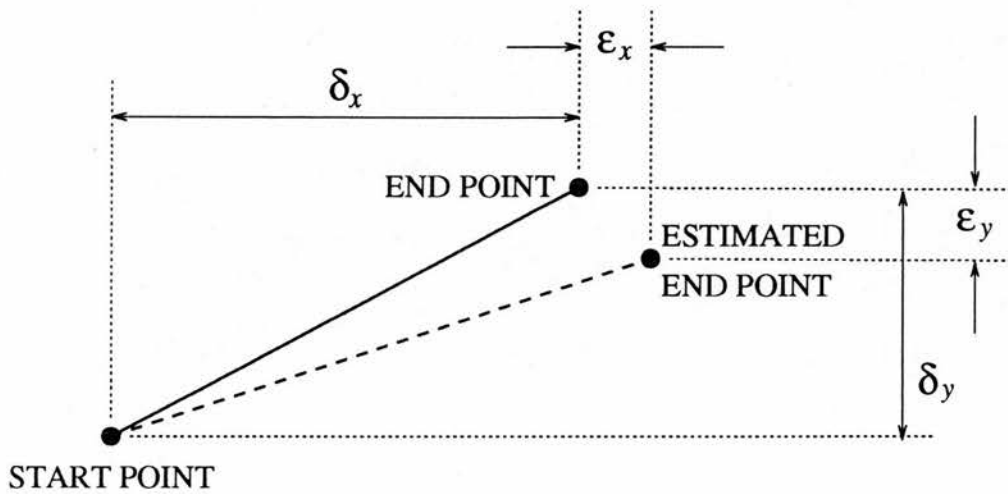


Figure 3.5: Error in animal's estimate of its movement: each time the animal moves (from start point to end point), its estimate of the distance and direction it has moved is slightly inaccurate.

### 3.4.1 Increasing Expected Error with Increasing Movement

The first navigational phenomenon to implement is that of the animal being able to get lost. The animal's knowledge of where it is becomes increasingly inaccurate as it moves further and further away from the den. At each timestep, if the animal moves a distance  $\delta_x$  in the  $x$ -direction and  $\delta_y$  in the  $y$ -direction then it is assumed to incur errors in its estimates of how far it moved of  $\epsilon_x$  and  $\epsilon_y$  (i.e. it assumes it has moved  $(\delta_x + \epsilon_x)$  in the  $x$ -direction and  $(\delta_y + \epsilon_y)$  in the  $y$ -direction, as shown in figure 3.5).  $\epsilon_x$  and  $\epsilon_y$  are both random numbers drawn from a gaussian distribution with mean 0.0 and variance  $\sigma^2$  ( $= 0.05$ ). Therefore after one movement the expected variance of the estimates of the  $x$  and  $y$  coordinates will be  $\sigma^2$ , and after  $N$  movements the variance will be  $N\sigma^2$ .

### 3.4.2 Deciding Whether to Recognise a Feature

So far the animal will move about the environment, gradually accumulating more and more error in its estimate of where it is, and only reducing the variance of this estimate when it manages to return to its den. In order to create a more realistic model of navigation, in which the animal will not tend to get lost in areas it is well acquainted with, it is necessary to allow the animal to *recognise features*. When it encounters a recognised feature (one it realises it has perceived before) then the memory of where it thought the feature was before will be used to improve its estimate of where it currently is. The way in which it is decided whether or not a feature is recognised is now described. The animal can only recognise a feature in the same square as itself.

In order for a perceived feature to be 'matched' to a remembered one (i.e. to be recognised, whether correctly or not), the following conditions must be true:

1. The perceived and remembered feature must be of the same type (e.g. both water sources).
2. The values of any properties of the features should be similar except for food and water source values which can be cyclic or dependent on the weather or changed by the animal consuming them.
3. The difference in the estimated positions should be small compared to the sum of the variances of the animal's position and the remembered position of the feature.
4. The sum of the two variances should not be excessive.

The third and fourth conditions are achieved by thresholding a term known as the **Mahalanobis distance** added to the logarithm of the combined variance. That is to say, a match is only possible if

$$\left( \frac{(x_A - x_F)^2 + (y_A - y_F)^2}{(\sigma_A^2 + \sigma_F^2)} \right) + \log(\sigma_A^2 + \sigma_F^2) < 5.0$$

where  $(x_A, y_A)$  = estimated animal coordinates,  $(x_F, y_F)$  = estimated feature coordinates,  $\sigma_A^2$  = variance of estimate of animal's own position, and  $\sigma_N^2$  = variance of estimate of feature's position (before this encounter). The figure of 5.0 was determined empirically. It gives a good balance between (i) not rejecting too many true matches and (ii) not accepting too many false matches.

When deciding which remembered feature, if any, is matched with the perceived one, the map is scanned in order of memory strengths (described below), with highest memory strengths being considered first. The first entry to satisfy all of the conditions above is matched with the perceived one, otherwise the perceived feature is not recognised and a new memory entry will be added to the map. This scheme allows a single feature to be represented in the memory more than once, and also allows two similar and adjacent features to be represented by the same memory. Recognition is distinctly error-prone.

**Landmarks** are easily distinguishable features which are not confused with one another. Because of this they are likely to have more reliable coordinates, and thus to make it easier for the animal to navigate successfully around the environment.

### 3.4.3 Updating the Animal and the Feature after an Encounter

This section describes what happens when an animal encounters a feature. The feature can be recognised or can be treated as being perceived for the first time. The changes in the animal's estimate of where it is and in the variance of that estimate will be considered, as well as the changes that are made to the estimate of the feature's position, and the variance of that estimate. Finally, the changes in the estimates of the values of the feature's properties are described.

**Estimating feature positions** — the animal remembers the positions of features it encounters. The estimate of a feature's position is likely to become increasingly accurate as the animal visits it more and more often. This is implemented in the following way:



1. when an animal first encounters a feature then the remembered position of the feature is set to the animal's estimate of its own position.
2. similarly, the variance of the estimate of the feature's position is set to the variance of the estimate of the animal's own position.
3. when an encountered feature is matched to a remembered one (i.e. is recognised) then the new estimate of the feature's position is calculated according to the following equations:

$$x_{N+1} = \left( \frac{\left[ \frac{x_N}{\sigma_N^2} + \frac{x_A}{\sigma_A^2} \right]}{\left[ \frac{1}{\sigma_N^2} + \frac{1}{\sigma_A^2} \right]} \right), \quad y_{N+1} = \left( \frac{\left[ \frac{y_N}{\sigma_N^2} + \frac{y_A}{\sigma_A^2} \right]}{\left[ \frac{1}{\sigma_N^2} + \frac{1}{\sigma_A^2} \right]} \right) \quad (3.1)$$

where  $(x_N, y_N)$  and  $(x_{N+1}, y_{N+1})$  are estimates of the feature's position after  $N$  and  $N + 1$  encounters,  $(x_A, y_A)$  is the estimate of the animal's position,  $\sigma_N^2$  is the variance of the  $N^{\text{th}}$  estimate of the feature's position and  $\sigma_A^2$  is the variance of the animal's estimate of its own position. This equation is known to produce a minimum-variance estimate of the new position, as long as  $(x_N, y_N)$  and  $(x_A, y_A)$  are independent.

4. when a previously encountered feature is recognised then the new variance of its estimated position is calculated according to the equation:

$$\sigma_{N+1}^2 = \left[ \frac{1}{\frac{1}{\sigma_N^2} + \frac{1}{\sigma_A^2}} \right] \quad (3.2)$$

Equation 3.1 above is structured so that if the animal's variance is small (its position is accurately known), but the feature's is high (inaccurately known), then a big change will be made to the estimated position of the feature, whereas if the opposite is true then the change in the estimated position of the feature will be negligible. Equation 3.2 is the correct way of calculating the variance of the new estimate of the feature's position when the new position is estimated as in equation 3.1.

**Estimating the animal's position** — a choice needs to be made as to how the animal's estimated position and variance will be changed upon encountering a recognised



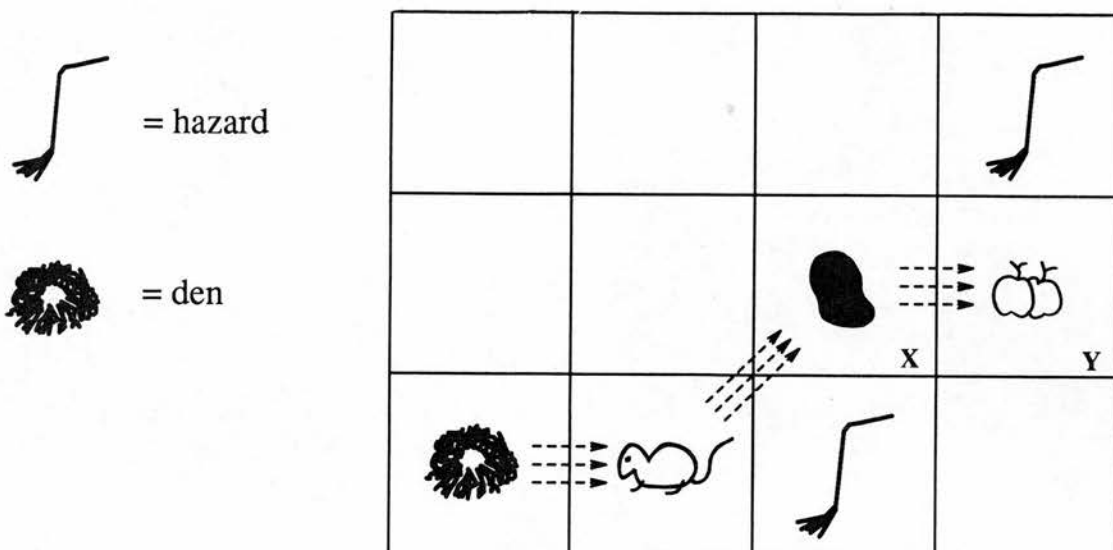


Figure 3.6: Travelling to one feature by way of another: square Y is usually visited by the animal just after it has visited square X, and so successive estimates of the position of the feature in square Y are not independent.

feature. The simplest and 'purest' choice is to set the animal's coordinates and variance to that of the feature it has just recognised, since those have already been changed to take account of the animal's estimated position in a principled way. This option also appears optimal in that it results in the minimum variance. There are two hidden difficulties with this though: (i) The animal will occasionally match the wrong remembered feature to the perceived one, and this will badly upset the animal's estimate of its own position, (ii) If the animal usually tends to travel to one feature by way of another (see figure 3.6), then the different estimates of the second feature's position will not be independent since they will always be affected by an error in the estimates of the first feature's position. If there is a large error in the initial calculation of the first feature's position then this will gradually die away with subsequent visits. If the animal simply sets its own position to that of the first feature each time it passes it, then the second feature's estimated position will only start to become accurate some time after the first one's does.

For these two reasons it is better to strike a balance between (i) not changing the animal's coordinates at all (and thus ensuring independent estimates) and (ii) changing directly to the feature's coordinates (thus taking full advantage of the visit to the feature and reducing the variance as much as possible). A tradeoff is made by setting the coordinates as follows:

$$x_{N+1} = \frac{1}{2}(x_N + x_F), \quad y_{N+1} = \frac{1}{2}(y_N + y_F)$$

where  $(x_N, y_N)$ ,  $(x_{N+1}, y_{N+1})$  are the animal's estimates of its own position before and after taking the feature into account, and  $(x_F, y_F)$  is the animal's estimate of the feature's position *after* the encounter. This tradeoff has been tested in the SE and shown to work well. Although it results in theoretically larger variances than simply adopting the features' coordinates (as in equation 3.1), in practice, because successive estimates are not usually independent, it results in more reliable navigation.

The variance of the animal's position is therefore calculated according to the equation:

$$\sigma_{N+1}^2 = \frac{1}{2}(\sigma_N^2 + \sigma_F^2)$$

where  $\sigma_N^2$ ,  $\sigma_{N+1}^2$  are the variances of the animal's estimate of its own position before and after taking the feature's position into account, and  $\sigma_F^2$  is the variance of the estimate of the feature's position.

**Estimating values of feature properties** — as with a feature's position, the animal is likely to get increasingly accurate estimates of a feature's properties the more times that it visits it. This is implemented as follows:

1. on the first encounter the estimates of the feature's property values are set to the perceived values.

2. on subsequent encounters they are set according to the following formula:

$$V_{N+1} = \frac{(V_N \times M_N) + (V_A \times (M_{N+1} - M_N))}{M_{N+1}}$$

where  $V_N$  and  $V_{N+1}$  are estimates of the value of the feature property before and after the current encounter,  $V_A$  is the animal's currently perceived value of the feature property and  $M_N$  and  $M_{N+1}$  are the strengths of the memory of the feature before and after the current encounter.

As explained below, the memory strength of a feature will decay with time if the feature is not re-visited, but will increase if the feature is visited and recognised frequently. The above formula makes successive estimates likely to be more and more accurate by weighting each new contribution by smaller and smaller amounts.

### 3.4.4 Memory Strengths and Forgetting

1. when a feature is encountered for the first time, or a previously encountered feature is not recognised as such, then the strength of the new memory ( $M$ ) is set to

$$M = U \times \frac{1.0}{\sigma_A^2}$$

where  $U$  = a measure of the 'utility' of the feature (e.g. its food value).

2. on subsequent recognitions the memory's strength is updated according to the equation

$$M_{N+1} = M_N + \left( U \times \frac{1.0}{\sigma_F^2} \right)$$

3. at every timestep each memory strength is slightly decreased. This is implemented by multiplying by a number slightly less than 1.0 (0.987). This makes the strength of each memory decay slowly over time if it is not encountered and recognised by the animal.

4. in addition, when the animal tries to use the map to find a feature, but does not find it where it expects to, then the strength of the feature's memory is decremented if the animal's position variance is small.

The high-level effect of these four rules is that those features which are frequently recognised, have been recently recognised, and have a high utility will have a high memory strength. Conversely, those feature memories which have not been recognised for a long time, or which are only very infrequently encountered or which have a low utility will have lower memory strengths.

The 'map' has a finite size (100 feature memories), and is ordered according to memory strength, with the highest at the top. When a feature is encountered but not recognised (is treated as novel) then its memory strength is calculated and it is slotted into the appropriate place in the map. If the map is already full then a new memory will result in the bottom one 'dropping out'. Over time old, redundant memories fall out of the map. This is important since (i) features can disappear from the SE, (ii) features a long way from the den can be visited just once, and (iii) multiple memory entries can initially be generated for the same feature (because it was not recognised on subsequent visits) of which one will eventually come to dominate.

### 3.4.5 Using the Map

When deciding which of the various remembered features of a particular type to visit (e.g. which remembered food source to head to), or whether to head to any of them at all, the animal calculates an 'attractiveness' for each remembered feature. This attractiveness takes into account the following factors: (i) distance of remembered position from animal, (ii) remembered utility of feature, (iii) strength of memory, (iv) variance of remembered position and (v) probability of the feature being toxic.

Navigating successfully towards a remembered feature is not automatically accomplished, once the animal has decided to try and do so. The animal may not find the feature where it expects to, due to inaccuracies in its estimates of its position and of the feature's position. There are also competing sub-problems whose demands may interrupt the animal's passage to the remembered feature and 'knock it off course' (e.g. avoiding predators, avoiding hazards). It is also possible for the animal to get caught in a *cul-de-sac* (e.g. formed by dangerous place features). A record of how long the animal has been trying to get to a feature is kept, and the attractiveness of the remembered feature is reduced if the animal has not reached it after a reasonable time. Eventually this will make it give up looking for that feature.

If the animal reaches a remembered feature but finds that its utility is very low (e.g. if a fruit type food feature is in the wrong part of its cycle) then the attractiveness of that feature will be reduced for a time, so as to stop the animal heading back to it in the near future.

A final use of the map is to record which food and water features might be toxic. When an animal becomes sick because of toxic food or water then it looks through the map at all the food and water features that have been consumed recently and increases the estimated 'toxicity' of each of the memories according to (i) how much toxicity it is now suffering from, and (ii) how recently the food or water source was visited. Those food or water sources which are assigned a high toxicity in the map are unlikely to be visited again.

### 3.5 Motor Control

The third of the processes to be modelled (after perception and navigation) is motor control. In section 3.2 a list of 14 low-level actions was given (CLEAN, EAT\_CF, EAT\_FF, POUNCE, DRINK, REST, FREEZE, MOVE\_FAST, LOOK, LOOK\_AROUND, SLEEP, COURT, MATE and MOVE). MOVE\_FAST, LOOK and MOVE can all occur

in any one of 8 different directions and so there are actually 35 alternative actions available to the animal.

The model of motor control is fairly simple. The execution of actions by the animal is error-prone. It will usually execute the action properly, but not always. The probability of the animal carrying out an action incorrectly is 1%, unless its health is below 0.1, in which case the chance is  $(100 \times ((0.1 - \text{animal.health}) / 0.1))\%$ . In addition, if *animal.health* is less than 0.1 then the animal is incapacitated and is unable to move fast, mate or pounce (the most strenuous actions).

The choice of action by the animal has several effects: (i) the amount of health the animal can recover in a single timestep is dependent on the action it is undertaking. Actions such as sleeping and resting allow a lot of recovery while actions such as mating, pouncing and moving fast allow hardly any recovery. (ii) the action of the animal effects how conspicuous it is (how likely to be perceived by predators). Actions such as sleeping, resting and especially freezing are very inconspicuous, whereas actions such as pouncing and moving fast are not. (iii) the animal's food, water and temperature internal variables are linked to the animal's action. The animal's food and water levels are decreased to a greater degree by strenuous actions such as moving fast, mating and pouncing. These also increase the animal's internal temperature. Other actions which require a lower level of exertion, such as sleeping, resting and looking in a direction, will reduce fat, carbohydrate, protein and water levels by less, and will also lead to a smaller increase in body temperature.

Taken together, the effect of the three points above is to assign different **costs** to the different actions the animal can undertake.

### 3.6 Analytical Tools

In order to be able to examine the performance of different ASMs in the SE, and to be able to diagnose what is wrong with them, various analytical tools are required. The components of the SE that help in the analysis and comparison of ASMs are:

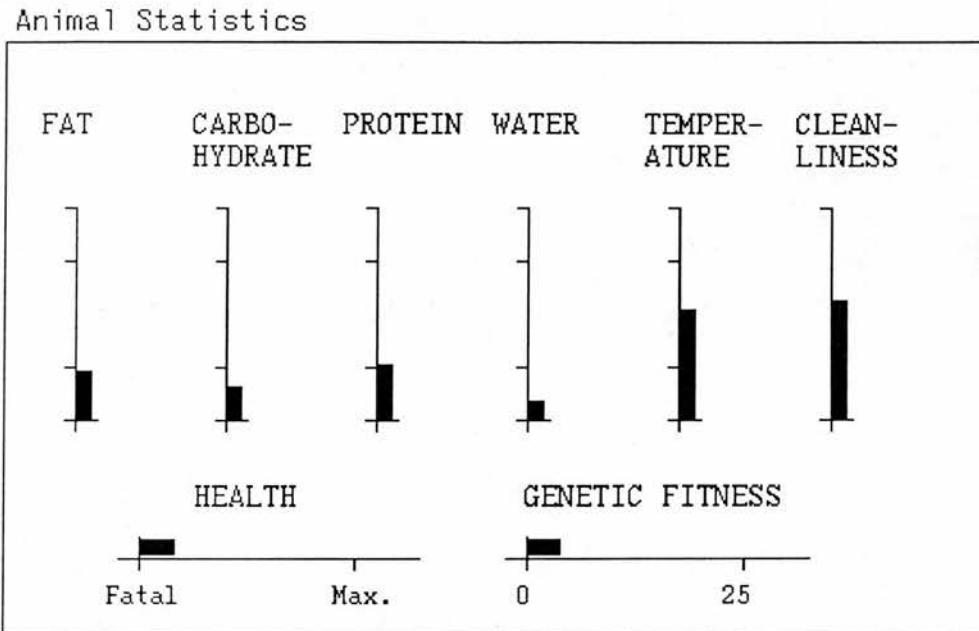


Figure 3.7: Display showing the values of the animal's internal variables and genetic fitness.

1. **graphical displays:** there is a full graphical display of the whole of the SE, together with displays showing the states of the animal's internal variables, what the animal perceives and the animal's navigational 'map'. These are shown in figures 3.7–3.10. These graphical displays allow the observer to see where in the environment the animal is, what perceptual inputs (internal and external) it is receiving, what navigational inputs it is receiving, and what actions the animal is choosing. Another display plots activations of different system-level entities through time, together with an indicator every time the animal performs an action relevant to that system (figure 3.11). By examining these different displays while an ASM is operating in the SE, any gross deficiencies in action selection can easily be detected.

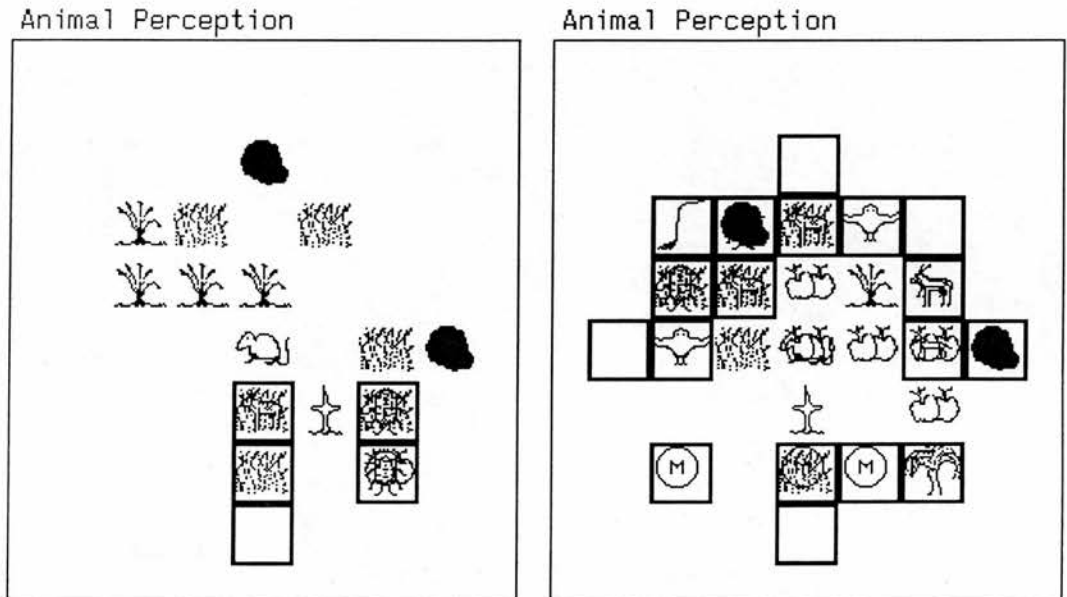


Figure 3.8: Display showing two instances of the animal's perception of its local environment. White boxes around squares indicate that the contents of the square have been incorrectly perceived. The perception on the left occurred when the animal was not in vegetation and chose the action LOOK\_AROUND. The instance on the right occurred when the animal chose the action EAT\_FF while in a square containing fruit type food.

2. **speed of SE:** The SE can be paused, stepped through successive timesteps, or made quicker or slower so as to give a clearer impression of what is going on.
3. **health file:** all events that affect the animal's health are written to a file. This file can then be examined at a later date to see what factors were most influential in the animal's death (e.g. to see if the main causes of death were lack of food, attacks by predators, old age or whatever).
4. **SE log file:** the initial setup of the SE, and all events (e.g. appearances of predators, changes in weather) that occurred during the animal's life are copied to a log file. The animal's perceptual and navigational inputs, choice of action



### Navigation

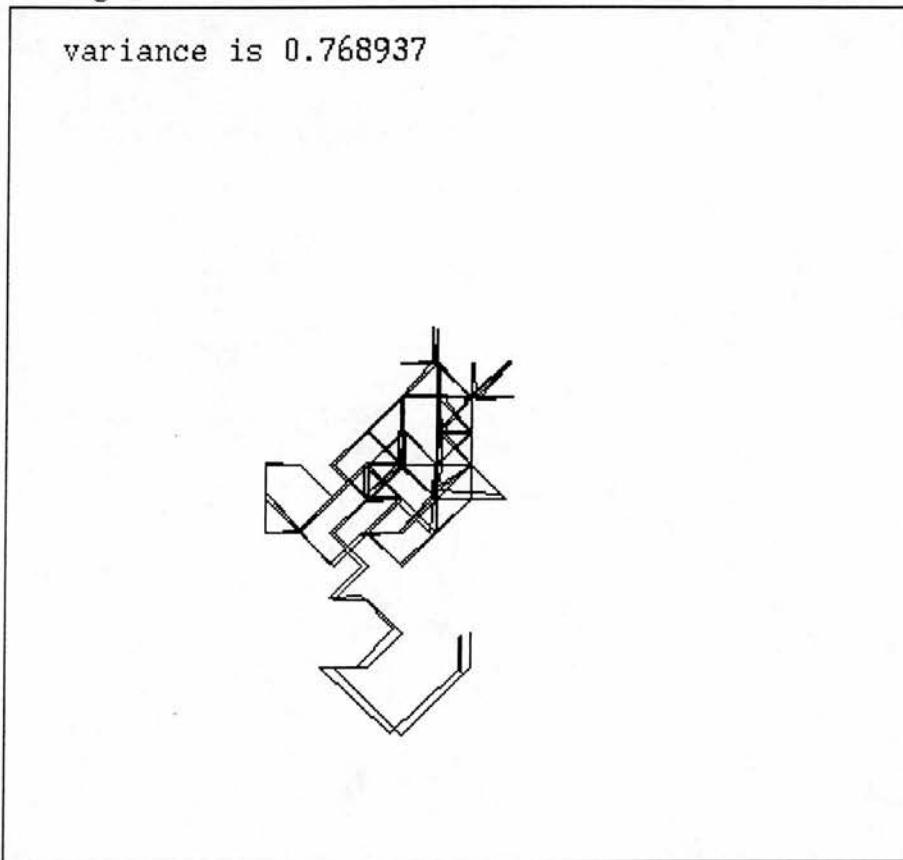


Figure 3.9: Display showing accrued error in estimate of position. The two lines represent the actual and estimated position of the animal. As the animal moves further away from known territory (the centre of the diagram), the animal's estimate of its position becomes progressively more and more inaccurate and the two lines diverge.

---

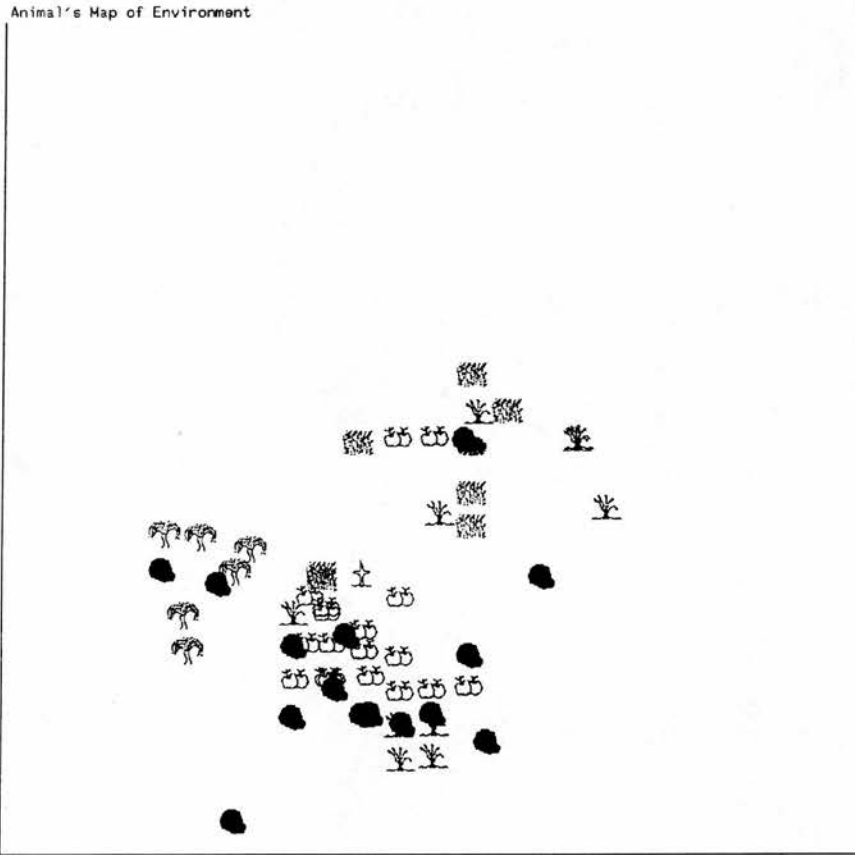


Figure 3.10: Display showing the animal's map of the environment after several hundred timesteps. Some features are not recognised on subsequent visits and so are represented more than once. The estimated position of some of the remembered features is not very accurate.

---

### Longitudinal Profiles

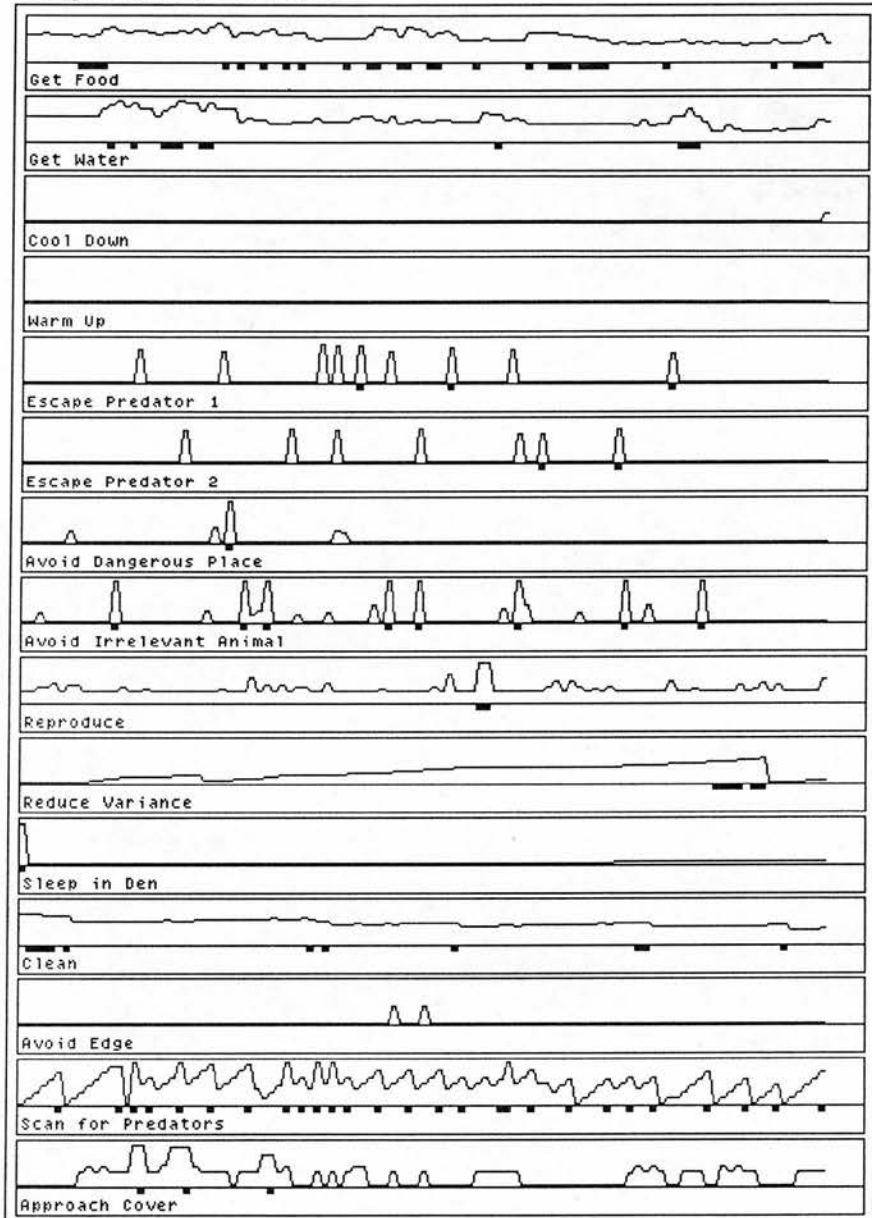


Figure 3.11: Display showing activation levels over time for each system in the drives ASM (after [Maes, 91]). The solid blobs beneath the axis of a system denote the selection of actions relevant to that system.

and success in executing that action are all written to the log file at each timestep. This allows a more detailed 'post-mortem' of what happened during the animal's life. This file can also be used to re-create the same SE for a different ASM, and to re-run it so that the same random events occur. This can be used to discover at what points different ASMs will select different actions if they are running in an identical SE.

5. **dbx**: the SE can be run within **dbx** (a debugging tool which allows all the variables, arrays, etc in a program to be examined during a program's execution) to give a more detailed step-by-step analysis of arbitrarily fine details of the operation of an ASM. For instance, *dbx* can be used to examine the excitations flowing along the links of a distributed ASM, and therefore can be used to discover *why exactly* an ASM is choosing inappropriate actions.

## Chapter 4

# Validation

Most of this chapter is devoted to describing the testing that was carried out on the SE. *Low-level* testing had the aim of ensuring that the mechanics of the SE worked as they were supposed to (e.g. a water source decreases in value when the animal drinks from it, predators perceive the animal less frequently when it is in vegetation). *High-level* testing had the aim of ensuring that the sub-problems were all individually manageable, and that the problem as a whole (satisfying all of the sub-problems simultaneously) was useful, i.e. that it produced measurably different levels of performance from ASMs of varying suitability. Even after all of the low-level mechanics of the SE were verified, the overall problem posed by the SE would not be interesting if either (i) the animal does not have enough time to keep the effects of all the sub-problems in check simultaneously, and is therefore going to die quickly whichever ASM is used, or (ii) the animal has so much time that it can deal with all of the sub-problems and still have lots of time to spare. In this case the choice of ASM will not lead to a large difference in performance.

The final part of this section derives equations to calculate the degrees of statistical certainty provided by different results.

## 4.1 Low-Level Tests

The low-level tests on each feature in the SE checked the following: (i) the initial number of features generated, (ii) the spatial distribution of the initial features, (iii) the generation of the feature property values, (iv) the disappearance of feature instances over time, and the appearance of new ones, (v) the dynamics of feature instances (e.g. how they move about the SE, how their property values change over time), and (vi) the interactions of the feature instances with the animal (how the instances affect the animal, and how the animal affects them).

The low-level tests also checked the effects of every animal action, both on the features of the SE, and on the animal's internal variables. Other dynamics of the internal variables were also examined. All the details of the models of perception, navigation and motor control were checked. Finally, the effects, or costs, of different animal actions in terms of conspicuousness, health recovery, efficacy of perception and internal variables were also verified.

It would be laborious and of little value to describe all the low-level tests carried out. Instead, two representative sets of tests are described. Other aspects of the SE were tested in an equally rigorous manner.

### 4.1.1 Example 1: Fruit Food Sources

The different low-level tests carried out on fruit food features were as follows:

1. there should be between 50 and 80 instances generated initially.
2. over 70% of these should be positioned randomly within 2 squares of a water source, and the rest should be positioned randomly anywhere in the SE.

3. the instances should be given random initial property values, within certain pre-specified ranges and with certain means.
4. about 10% of the initial instances should be toxic.
5. the value of the food at a fruit food source should change in a ripening-decaying cycle.
6. very occasionally an instance should disappear from the environment.
7. very occasionally a new instance should appear in the environment.
8. when the animal eats from an instance, the food value of that fruit food source should decrease appropriately, and the fat, carbohydrate and protein internal variables of the animal should increase appropriately.
9. after being eaten by the animal, the food value of a fruit food source should recuperate slowly.
10. an instance of fruit food should provide shelter to the animal (protection against attack and perception by predators and amelioration of extreme temperatures).
11. when the animal eats a toxic instance of fruit type food then it should suffer 0.0–0.4 decrement to health within the next 500 timesteps.

This is an example of the many low-level tests that had to be carried out on each feature of the SE.

#### **4.1.2 Example 2: Motor Control**

The low-level tests carried out on the model of motor control were:

1. the effects of the animal's actions on the SE were mostly covered in the tests on the interactions between features and the animal, but the animal's actions also affect the animal's internal variables and these interactions were checked (e.g. between the action CLEANING and the internal variable 'cleanliness').
2. the animal should normally execute about 99% of actions correctly.
3. when the animal's health is below 0.1 then this percentage should decrease.
4. when the animal's health is below 0.1 then the animal should not be able to mate, move fast or pounce.

This is an example of the low-level tests carried out on one of the three processes modelled. Perception and navigation are more complex and therefore required more extensive testing in the SE.

## 4.2 High-Level Tests

Whereas the low-level tests were carried out mainly to check the detailed workings of the SE, the high-level tests described in this section had the purpose of (i) checking that the individual sub-problems were achievable, and (ii) checking that the overall problem (dealing with all the sub-problems simultaneously) was set at a suitable level of difficulty. 'Achievable' is used in this case to imply that the effect of a sub-problem on the animal's health can be kept to a small amount with the allocation of sufficient time, at least for the majority of random set-ups of the SE (some SEs will be generated with, for instance, water instances which are all positioned a long way from the den, and so it will be impossible for the animal to get enough water). A 'suitable level of difficulty' above is such that different ASMs will give different performance results in the SE. The amount of time available to the animal, in comparison to the sum of the amounts of time required by each of the sub-problems, must be neither too great (AS problem too easy) or too small (AS problem too hard).



### 4.2.1 Sub-Problems

Each sub-problem is now examined in turn, and graphs are presented to verify that each one is solvable. In order to test that each sub-problem is individually solvable or manageable (if enough time is allocated to it), it is necessary to use an altered version of the SE and an altered ASM. The SE needs to be altered so that the only factors that impinge on the animal's health are those due to the sub-problem in question. The ASM needs to be altered because the amount of time being spent on the particular sub-problem needs to be able to be manipulated in some way.

The tests on the manageability of each sub-problem were therefore carried out by looking at the performance of an altered version of a simple but reasonably competent ASM (the 'drives' ASM) in a version of the SE in which the factors affecting health are all removed, except for old age and the factors relevant to the particular sub-problem being studied (e.g. decrement due to lack of cleanliness for the 'cleaning' sub-problem). In a drives mechanism (see section 8.1) a number between 0.0 and 1.0 (a drive strength) is calculated for each sub-problem. In theory each drive strength will represent how beneficial it would be to the animal to attend to that sub-problem, although in practice the drive strength will only be an approximation of this. The sub-problem with the highest drive strength is then selected, and the most appropriate action for that sub-problem is then calculated and executed. In the drives mechanism there is a competition between the sub-problems to capture the animal's attention and the closer the drive strength of any sub-problem to 1.0, the more likely it is to win the competition and get to choose the executed action.

The two graphs for each sub-problem are both derived using an artificially altered drive-type ASM in which the 'competition amongst drive strengths' is tampered with so that the particular sub-problem being studied will always be made to win the competition if its drive strength has a value  $> T$ , but will be excluded from the competition otherwise (see figure 4.1). The threshold  $T$  is given a random value between 0.0 (sub-problem always wins) and 1.0 (sub-problem never wins) at the beginning of each

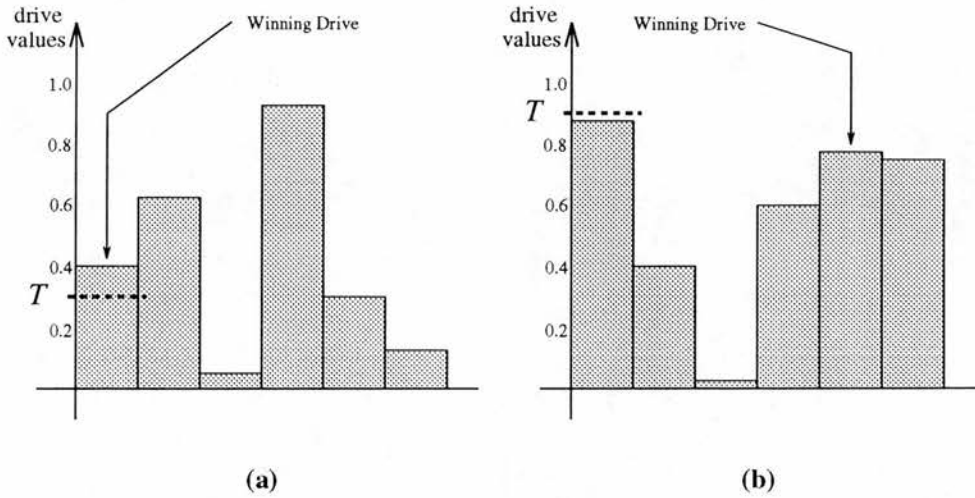


Figure 4.1: Artificial drive selection: **(a)** the drive for the sub-problem being investigated (the first bar in both graphs) wins the competition if it is greater than  $T$  (regardless of the values of the other drives), otherwise **(b)** a normal competition amongst drives (with the exclusion of that which is being investigated) takes place when the drive for the sub-problem being investigated is less than  $T$ .

test. Altering the value of  $T$  affects how much of the animal's time is spent on that sub-problem. If a sub-problem is achievable, then with  $T = 0.0$  the animal will die, on average, at an age very close to its maximum lifespan (since it will devote all its time to dealing with the only factors that can shorten its lifespan). There will be some reduction in average lifespan as  $T$  increases from 0.0 to 1.0 (when it is not devoting any time to ameliorating the factors reducing its health).

The first graph in the pair for each sub-problem plots the animal's lifespan against the value of  $T$ . 600 tests are carried out and each result is plotted as a point in the graph. In addition to plotting each result, averages are also plotted to give an idea of the trend as  $T$  changes from 0.0 to 1.0. To obtain these averages, results for different parts of the

range of  $T$  were grouped together. Each sub-division is of size 0.1, and so there are 10 sub-divisions, each with an average of 60 results. For instance, to get the first point in each curve, all the results for which  $T$  is in the range 0.0–0.1 are grouped together, and the average and standard deviation of the group of about 60 results is calculated. The solid circles on the graphs represent the average lifespan for that sub-division, and the asterisks are placed at one standard deviation above and below the average. The second graph in each pair uses the same set of 600 tests, but plots instead the animal health after 500 timesteps. Again, the circles and asterisks are used to plot means and standard deviations for each part of the range of  $T$ .

The “%time” figures above each average in the lifespan graphs show the average proportion of its time that the animal spent on the sub-problem in question for that range of values of  $T$ . The animal typically spends less and less of its time on the particular sub-problem as  $T$  increases from 0.0 to 1.0, as is to be expected.

**Cleaning**: every time the animal cleans itself it reduces the difference between current and maximum cleanliness by a factor of 0.15. Its cleanliness decreases by a small amount (about 0.001) each timestep. Even at  $cleanliness = 0.0$ , the decrement to health is only 0.5.

The graphs in figure 4.2 show that the cleaning sub-problem is easily manageable (partly because uncleanliness can never reduce the animal’s health by more than 0.5), and that the average effect on health varies with the amount of time spent on the sub-problem. Figure 4.2a in particular shows that if the animal spends about 2% of its time cleaning then the average health decrement is about 0.2, whereas if it spends about 5% of its time cleaning then the average health decrement is about 0.1. The drive strength for the cleaning sub-problem has a maximum value of 0.75, and so the sub-problem is never attended to when  $T$  is greater than 0.75. This explains the kink in the two graphs. The gradual decrease in health due to old age means that the animal always dies at about age 7.5 days if it spends no time on cleaning (when  $T$  is greater than 0.75).

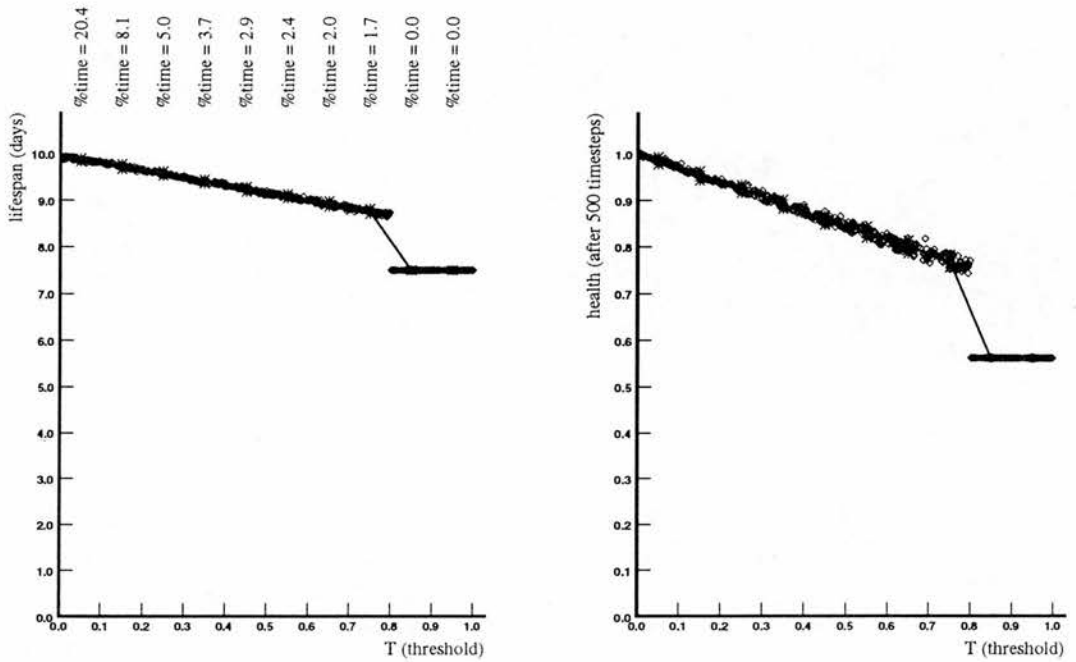


Figure 4.2: Test on cleaning sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

**Obtaining Food :** at every timestep the animal’s fat, carbohydrate and protein levels decrease by averages of about 0.0008, 0.0010 and 0.0005 (dependent on the animal’s chosen action). The animal can eat a maximum of 0.4 and 0.2 units of food per timestep respectively from fruit type or cereal type food sources. This corresponds to a maximum of 0.05 fat, 0.11 carbohydrate and 0.04 protein per timestep from a cereal food source and a maximum of 0.14 fat, 0.19 carbohydrate and 0.07 protein per timestep from a fruit food source. Prey instances vary randomly between 0.0 and 0.25 in food value (= 0.0–0.10 fat, 0.0–0.07 carbohydrate and 0.0–0.07 protein), which can all be eaten in a single timestep.

An average of 60 cereal food sources, 65 fruit food sources and 20 prey instances will be placed in the SE initially. An average of 30 instances per day of prey will traverse

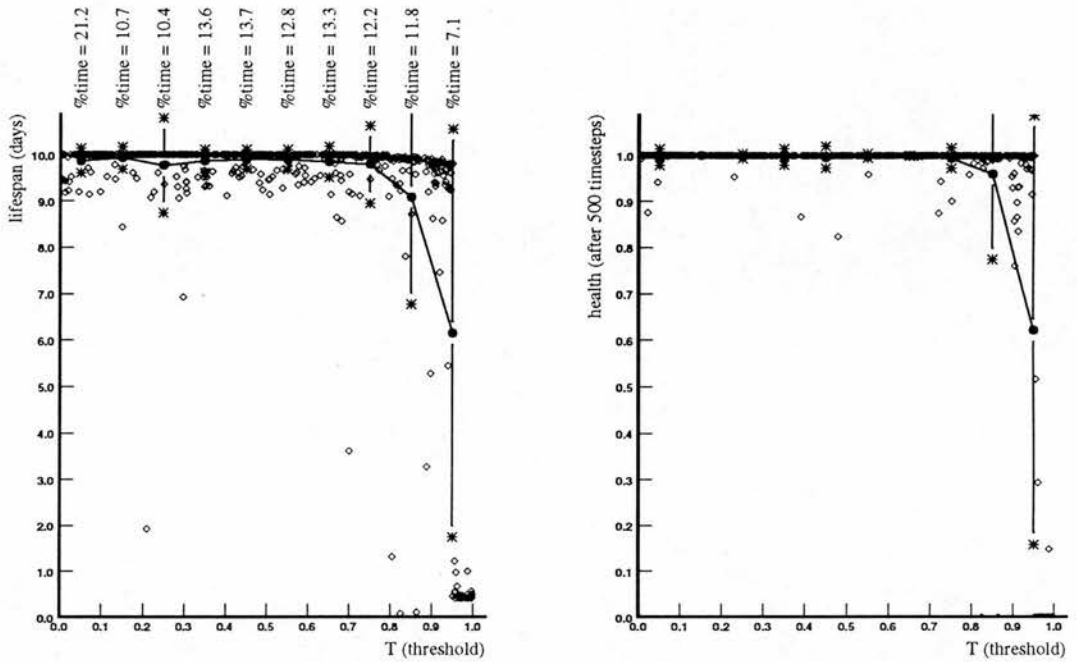


Figure 4.3: Test on obtaining food sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

the SE on a semi-random path.

The graphs in figure 4.3 show that obtaining food is easily manageable if the animal is able to allocate about 10–13% of its time to the sub-problem. It should be noted that the time being spent on the sub-problem does not change significantly as the threshold  $T$  increases from 0.1 to 0.9. This is because it takes as much food (and therefore time) to keep the food deficit constant at a lower level as it does to keep it constant at a higher level.

**Obtaining Water** : at every timestep the animal’s water level decreases by an average of about 0.0012 (dependent on the animal’s chosen action). The animal can drink a maximum of 0.06 units of water per timestep from a water source, which will contain

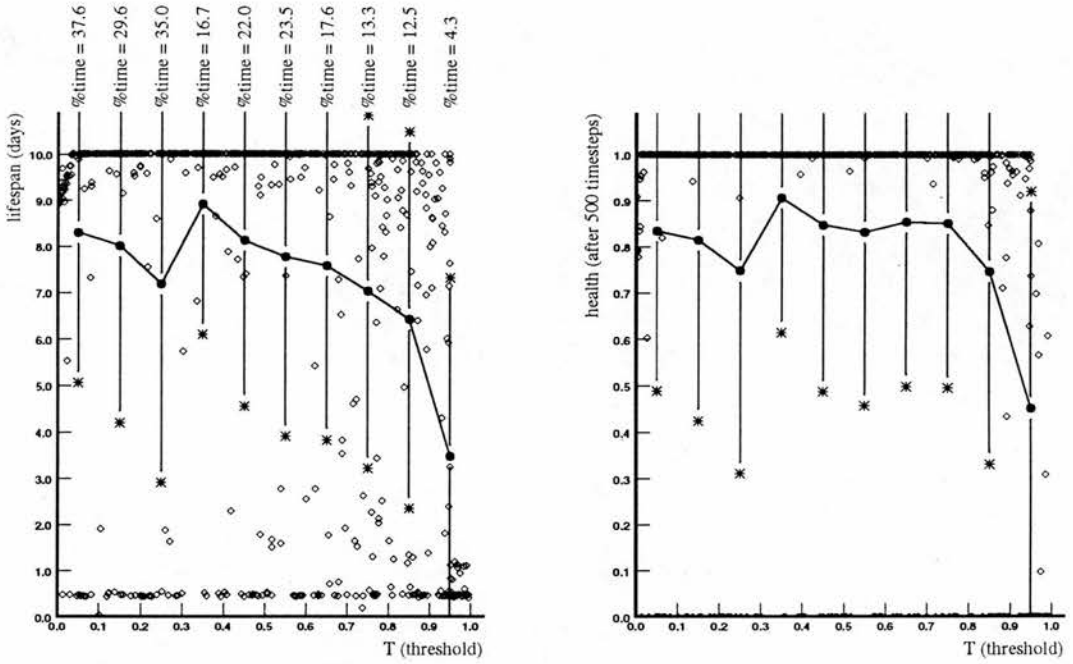


Figure 4.4: Test on obtaining water sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

an average of about 0.7 units of water. An average of 30 water sources are placed in the SE at the beginning.

The graphs in figure 4.4 show that the obtaining water sub-problem is to some extent manageable with the allocation of sufficient time, although there are always some early deaths whatever the value of  $T$ . This is due to the fact that certain random setups of the SE will contain a relatively small number of water sources, most of which are positioned far from the den. This will make it very difficult for the animal to get enough water. This phenomena is not so common with food because there are more instances of the three different types of food. This sub-problem requires an average of 15% or more of the animal's time. As with the graphs for obtaining food, the amount

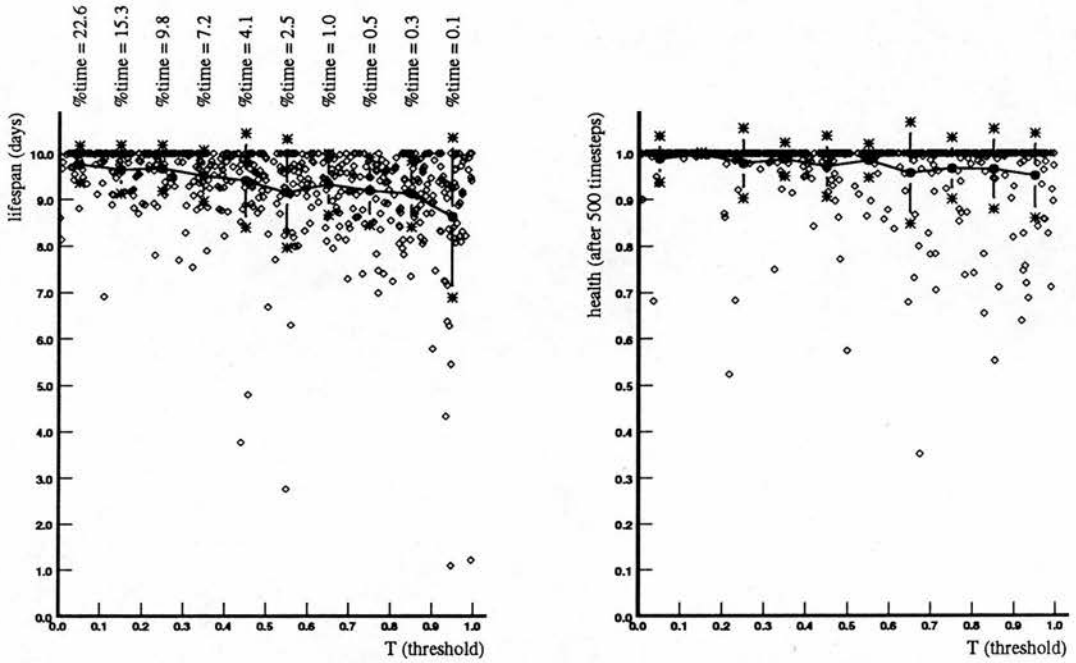


Figure 4.5: Test on temperature regulation sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

the external temperature, the animal's actions and the presence or absence of vegetation. The animal can take behavioural measures to reduce the effects of too high or too low temperatures by choosing a strenuous or non-strenuous action or by positioning itself in shelter or shade.

The graphs in figure 4.5 show that the temperature regulation sub-problem is manageable and that the effect on the animal's health varies according to how much time the animal devotes to the sub-problem. The animal needs to allocate about 5% of its time to this sub-problem.

**Predator Avoidance** : an average of 2 predator1s and 2 predator2s are placed in the SE at the beginning, and an average of 4 predator1s and 5 predator2s enter the SE every

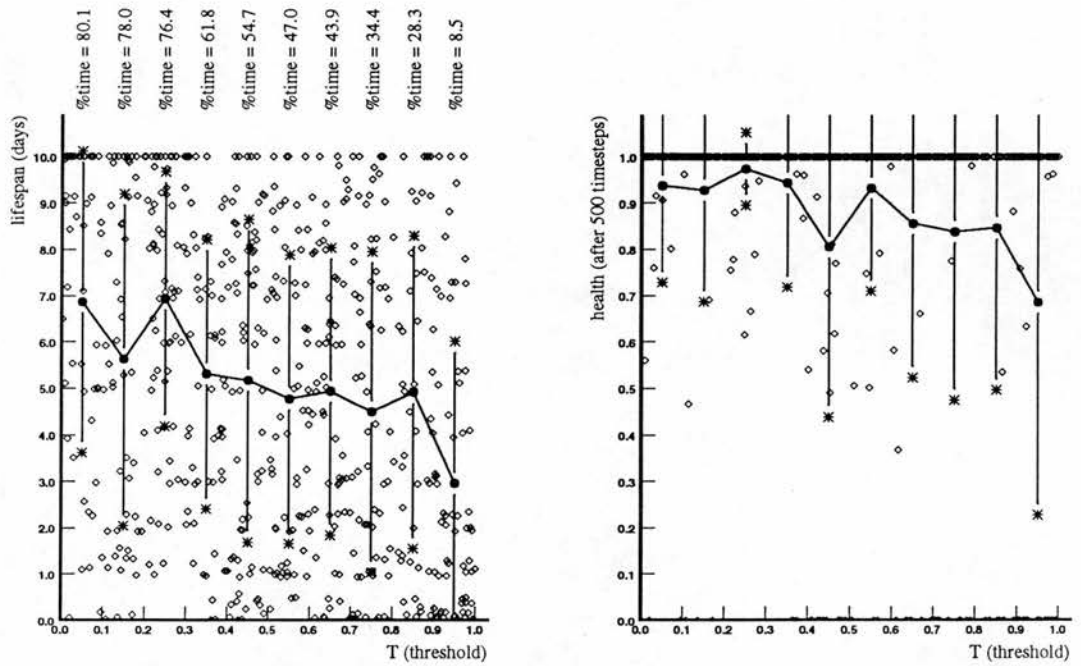


Figure 4.6: Test on predator avoidance, vigilance and staying close to cover sub-problems — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

day. If the animal perceives one in time it can try and run away or it can try and hide in vegetation. It can also try and freeze if it is in vegetation (this will reduce the chance of the predator seeing it and/or attacking it successfully).

The three different sub-problems of avoiding predators, vigilance and staying close to cover are all linked by a single type of decrement to the animal's health, that from attacks by predators. For this reason, the three sub-problems are examined together using a single test. In this case the only decrement to health comes from predator attacks *during the day* (any part of the day except night, when the animal cannot see predators to avoid them). The reason for excluding predator attacks during the night is that the probability of these attacks is not greatly affected by the time spent on these



three sub-problems (since the animal cannot perceive anything at night). The artificial drive selection is altered so that if any of the drives for the three sub-problems is greater than  $T$ , then the largest of the three is made to win the competition, regardless of the values of any other drives. If all three drives strengths are less than  $T$ , then the largest other drive wins the competition.

The animal is always prone to being attacked and killed by predators whatever its selection of actions (i.e. the task is not completely manageable even if 100% of the animal's time is spent on the three sub-problems), but the probability of health decrements due to predators can be reduced by allocating more time to avoiding predators, to vigilance behaviour, and to staying close to cover (as shown in the graphs in figure 4.6). Note that the animal does not benefit greatly from spending more than 30% or so of its time on these three sub-problems, unless it progresses to spending over 75% of its time, which it would be unable to afford.

**Vigilance** : the animal can see one square further and has a much better chance of perceiving squares correctly when it chooses the action LOOKING\_AROUND. Deciding to LOOK in a certain direction means that the animal has an even better chance of perceiving squares correctly in that direction. This sub-problem was tested together with predator avoidance and staying close to cover. The results are shown in figure 4.6.

**Staying Close To Cover** : the animal can reduce the risk of predation by staying close to its den or instances of cover, shade or fruit-type food (which all decrease the chance of a predator perceiving the animal and successfully attacking it). This sub-problem was tested together with predator avoidance and scanning for predators. The results are shown in figure 4.6.

**Edge Avoidance** : The animal just needs to avoid the edges of the SE. The graphs in figure 4.7 show that this sub-problem is hardly ever a factor affecting the animal's health. When it is then it is easily manageable with the allocation of a small amount of the animal's time (0–5%).

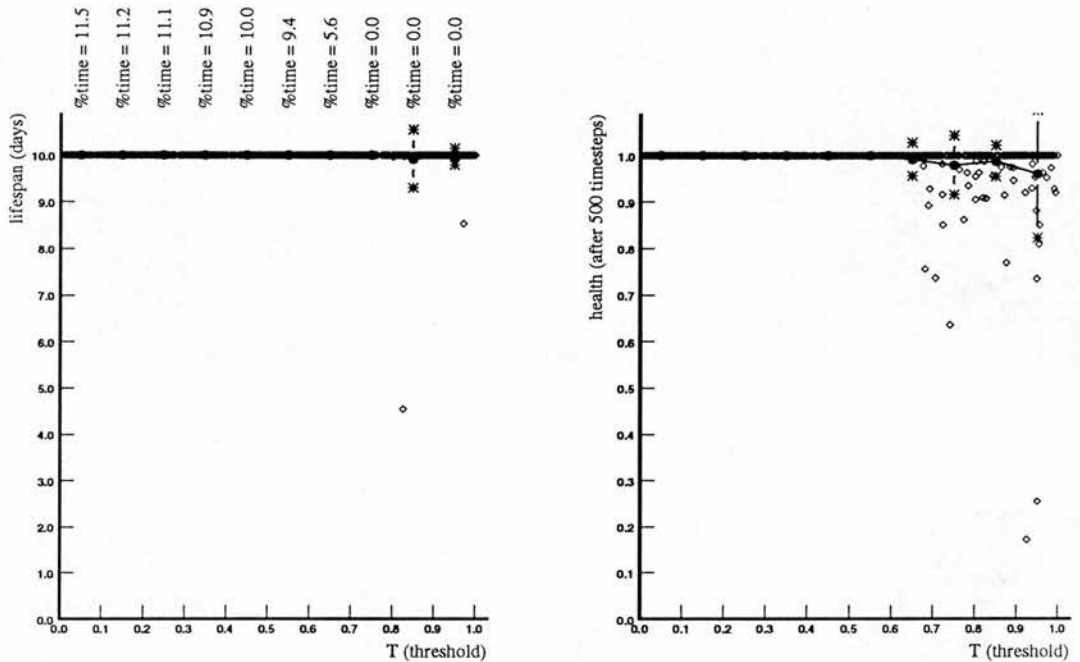


Figure 4.7: Test on edge avoidance sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

---

**Hazard Avoidance :** an average of 13 dangerous place features are placed in the SE when it is set up. The animal needs only to avoid these. This sub-problem is fairly manageable if the animal is able to spend 5–6% of its time on it (as shown in figure 4.8). In this test decrements to health were only considered if they occurred during the day since the animal cannot see dangerous places at night and so cannot avoid them.

**Irrelevant Animal Avoidance :** An average of 7.5 of these features traverse the SE every day. An average of 13 are placed in the SE when it is set up. The animal only needs to avoid them.

The graphs for this sub-problem (figure 4.9) were also produced by considering only

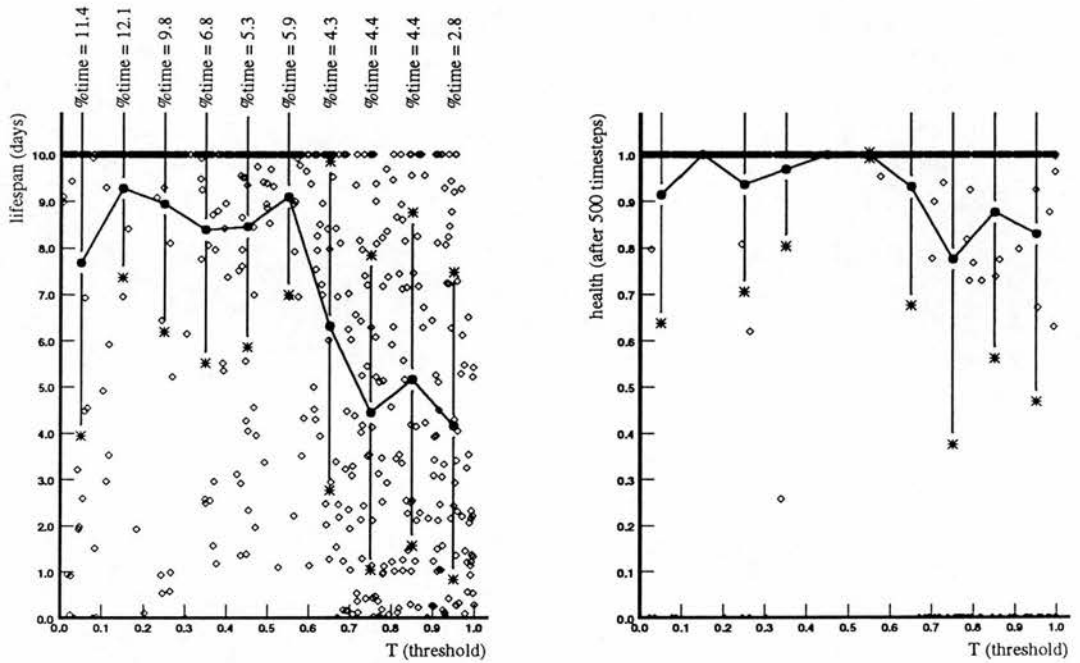


Figure 4.8: Test on hazard avoidance sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

decrements to health which occurred during the day. The graphs show that the sub-problem is manageable with about 5% of the animal's time.

**Sleeping at Night** : at night (which occupies  $\frac{1}{3}$  of the total time in the SE) the animal needs to remain in its den in order to avoid attack by the predators (the animal cannot perceive anything at night, whereas the predators can). Since the animal is unable to perceive dangerous places and irrelevant animals at night, it is also prone to injury from encountering them, should it decide to move about at night.

To produce the graphs in figure 4.10, the only decrements to health were from old age, predators attacking at night and from encounters with dangerous places and irrelevant animals at night. The graphs show that the problem is manageable as long as the animal

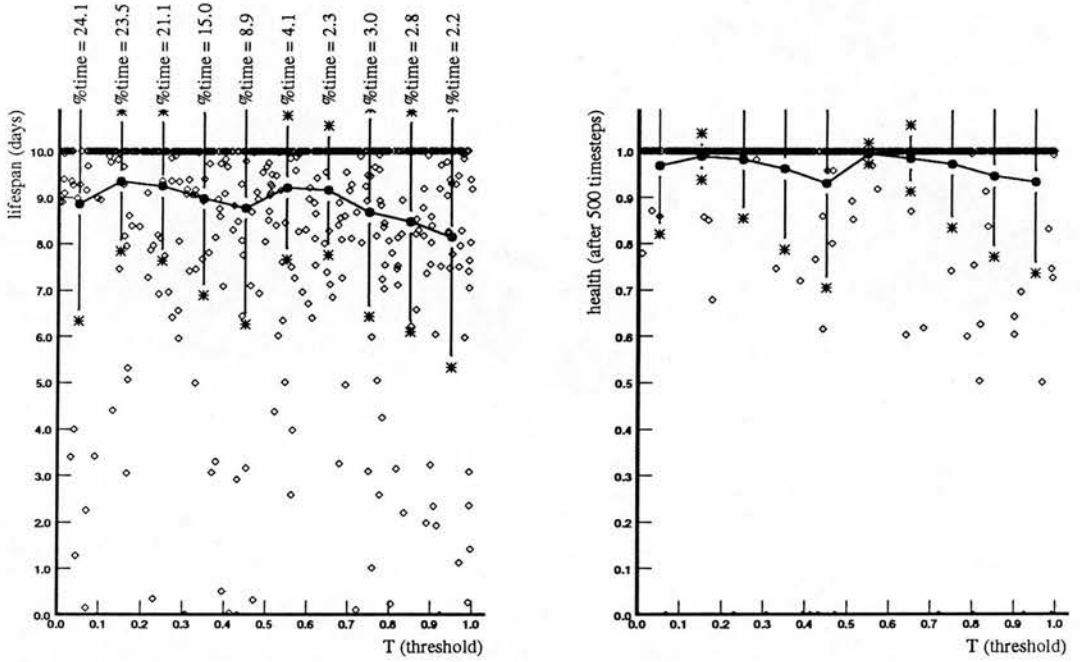


Figure 4.9: Test on irrelevant animal avoidance sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

does indeed spend the whole night (i.e. one third of its time) sleeping, as well as a short amount of time to get back to the den from wherever it is before nightfall.

**Not Getting Lost :** this sub-problem is only indirectly important in relation to being able to find the den at night and heading to remembered food and water. When this sub-problem is chosen then the animal heads back towards the den. It will either encounter remembered features on its way back towards the den (and so probably improve the accuracy of where it thinks it is) or else it will eventually get back to the den, where its variance is reduced to zero.

For the purposes of producing the graphs in figure 4.11, the only decrements to health considered were those used for testing the sleeping sub-problem (predators, dangerous

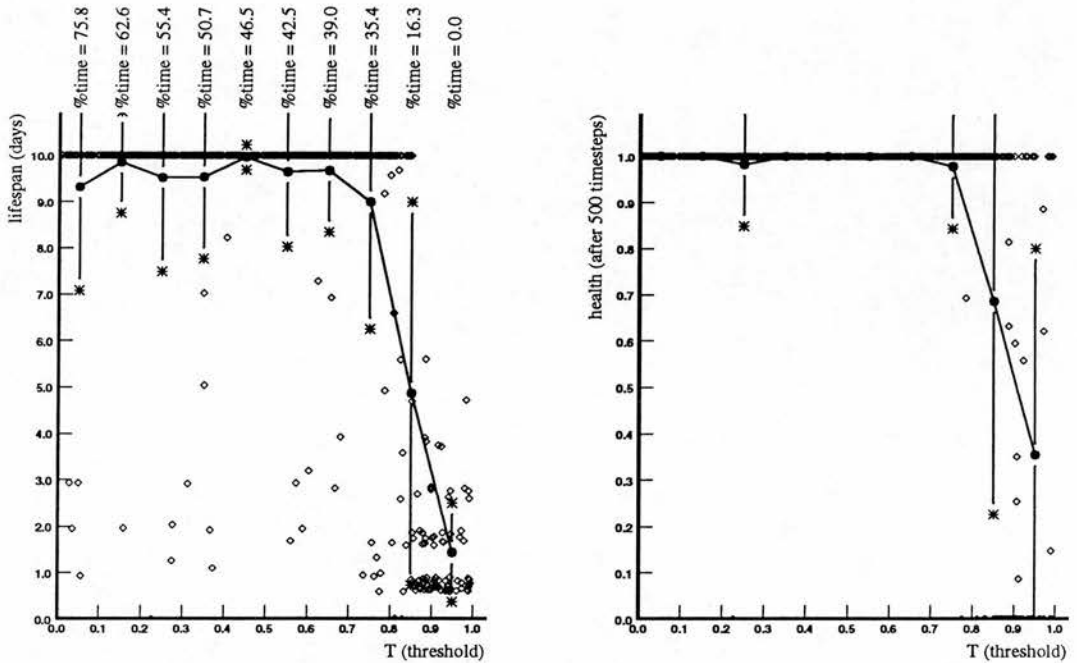


Figure 4.10: Test on sleeping at night sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

places and irrelevant animals at night), and also decrements due to shortages of food and water. The results show that it is very important for the animal to spend some time on this activity (4% or more), although not too much. Unlike other graphs, average health and lifespan decrease as a lot of time is spent on this sub-problem. This is because spending a lot of time on this sub-problem inhibits exploration and hence makes it harder for the animal to find enough food and water (as well as reducing the time available for getting food and water).

**Reproduction** : An average of 30 potential mates traverse the SE every day. The animal needs to court a receptive mate and then mate with it in order to increase its genetic fitness by 1. Attention to this sub-problem does not usually affect health and

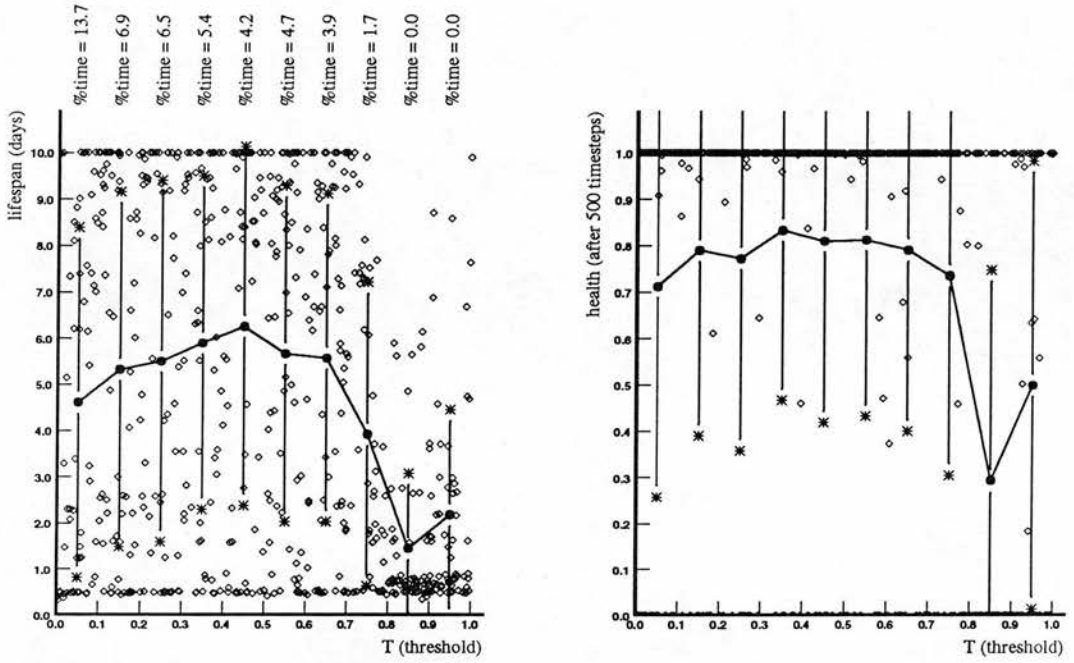


Figure 4.11: Test on not getting lost sub-problem — variation in lifespan and health after 500 timesteps for an artificially altered drive-type ASM. Further explanation in text.

survival (although the animal can be attacked by a mate). There is only an indirect effect on health, because time spent trying to reproduce detracts from time available to concentrate on survival.

To produce the graphs in figure 4.12, all decrements to animal health were allowed. 1500 tests were used instead of the usual 600. The ASM was altered in a similar manner to that of other tests. The sub-problem 'reproduction' won the competition if its drive strength was greater than  $T$ , and was removed from the competition otherwise. The left hand graph of figure 4.12 plots lifespan against  $T$  as normal. The right hand graph plots genetic fitness against  $T$  (instead of health after 500 timesteps). The graphs show that the animal has to strike a balance between spending time on survival and spending

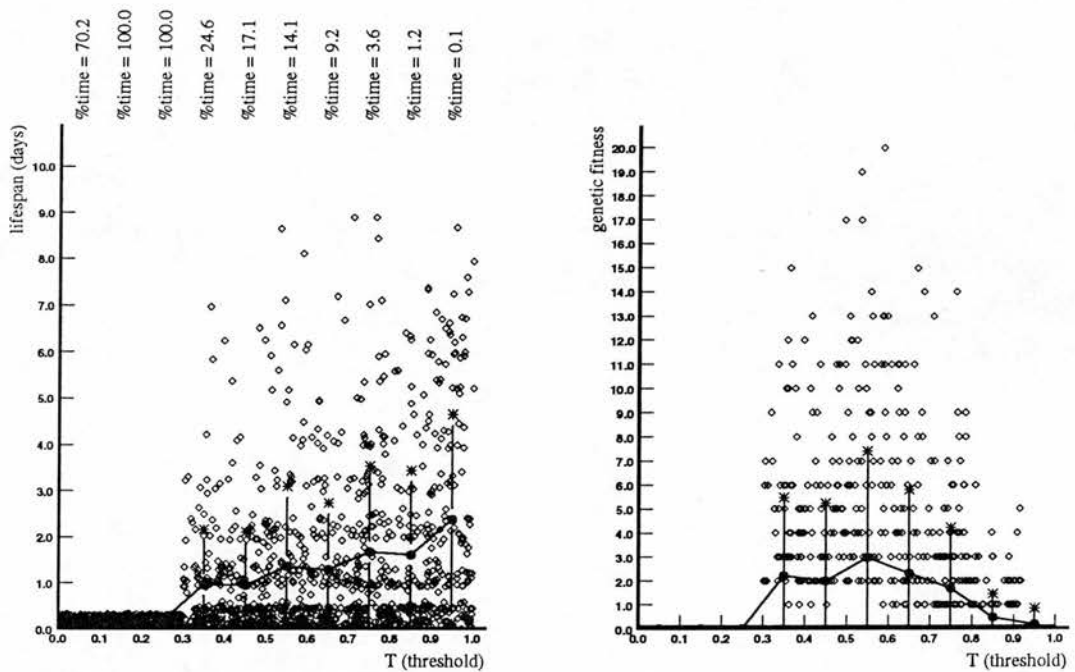


Figure 4.12: Test on reproduction sub-problem — variation in lifespan and genetic fitness for an artificially altered drive-type ASM. Further explanation in text.

time looking for and copulating with mates. In the SE, with this particular ASM, the optimal trade-off occurs when approximately 14% of the animal's time is spent trying to reproduce. More than this and the animal dies at a young age (and hence has none, or only a few chances to increase its genetic fitness). Less than this and the animal survives for a longer time but doesn't attach enough importance to reproduction and hence mates only very infrequently, if at all.

The left-hand graph shows that most animals die during the day time (the animal is started off in the SE at the beginning of the day, i.e. at sunrise). This is why the likelihood of death varies periodically. More animals die during the daylight part of the day because the animal will usually sleep in its den during the whole of the night.

### 4.2.2 Tests on Combined Problem

An indication of how difficult it is to survive in the SE can be obtained by summing up the average amounts of time that each sub-problem, apart from reproduction, requires. So for instance, it was noted that 'cleaning' required at least 2% of the animal's time to keep the effect on the animal's health to a sufficiently small decrement, and 'getting food' required at least 10%. Summing all of these amounts gives a total of  $(2 + 10 + 15 + 5 + 30 + 0 + 5 + 5 + 33 + 4) = 109\%$ . Since figure 4.12 indicates that it is optimal to spend about 14% of the animal's time reproducing, the drive-type ASM will only have about 86% or so of time to distribute among sub-problems collectively requiring 109% or more.

A second test on the whole problem for an ASM is to check that most instances of a competent ASM do not just die off immediately. The graph in figure 4.13 shows the distribution of survival times for instances of the unaltered drive ASM (unaltered except that reproduction is prohibited from winning the competition and so the animal's only concern is survival), and shows that the problem is reasonable since a significant proportion of instances manage to survive for more than one day. The graph was calculated from the results of 1200 tests.

The third and most important test is to see if the SE will give different performance results for ASMs which obviously vary in the inherent plausibility of their action selection. The whole purpose of the SE is to provide a means of evaluating action selection mechanisms, and so the performance measurements that are produced need to reflect the difference between good and bad action selection mechanisms.

Eleven drive-type ASMs were tested (this time with reproduction *not* prohibited). The first ASM was a completely normal drive-type ASM with straightforward 'competition between drive strengths' to decide which sub-problem was most appropriate for paying attention to. The other ten drive-type ASMs also contained competition between all drive strengths, but the drive strengths were interfered with by adding noise to them



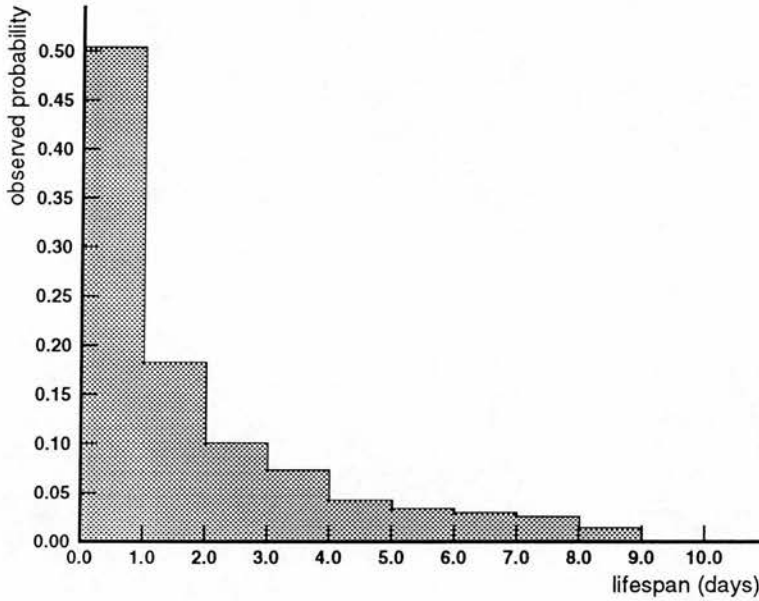


Figure 4.13: Survival times for the drive-type ASM. Many instances survive for a reasonable time, i.e. longer than a day.

---

just before performing the competition. In each case the equation below was used to calculate the noisy drive strength ( $D_n$ ) for each sub-problem:

$$D_n = D_0 + G(0.0, K)$$

where  $D_0$  is the undistorted value of the drive strength and  $G(0.0, K)$  is a gaussian random variable with mean 0.0 and standard deviation  $K$ . For the second ASM  $K$  was given the value 0.05, and for the other nine ASMs  $K$  was given the values 0.1, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45 and 0.50. Increasing values of  $K$  interfered more

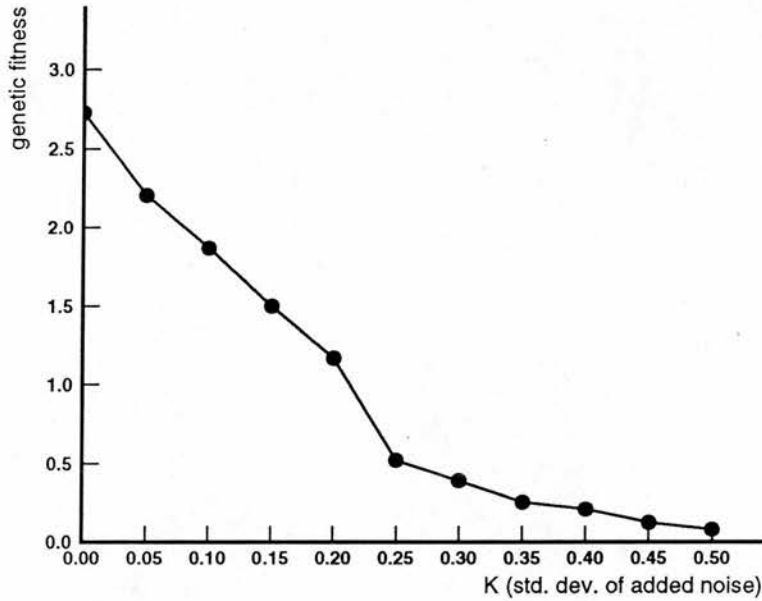


Figure 4.14: Performance (as measured by the SE) of drive-type ASMs distorted by varying amounts of noise added to the drive strengths. Further explanation in text.

---

and more with the selection of the most appropriate sub-problem and so it was to be expected that the first test would give the best result, followed by the second, third, fourth and so on. The resulting average performances are shown in figure 4.14, and it can be seen that the SE is indeed able to pick up differences in the suitability of the choices of actions made by these different variants of the drives ASM. Each point in the graph was obtained by averaging the results of 300 tests with that value of  $K$ .

### 4.3 Obtaining Statistically Significant Results

The first value to calculate is the approximate number of tests that need to be carried out in order to get a reliable measure of the performance of an ASM. A *reliable measure* will in this case be defined as a 95% certainty that the population average genetic fitness lies within  $\pm 5\%$  of the measured average from a sample. That is to say, the number of tests to be carried out ( $n$ ) will be enough to provide a 95% probability of the real average lying in the range  $(\bar{x} - 0.05\bar{x})$  to  $(\bar{x} + 0.05\bar{x})$ , where  $\bar{x}$  is the measured average from a sample of size  $n$ .

In order to obtain this degree of accuracy, the true variance of the sample averages  $\sigma_{\mu}^2$  needs to satisfy the inequality

$$1.96 \times \sigma_{\mu} < 0.05\mu, \quad \text{or} \quad \sigma_{\mu} < 0.0255\mu$$

where  $\mu$  is the mean of the population. The true variance of averages from samples of  $n$  measurements ( $\sigma_{\mu}^2$ ) is related to the variance of individual measurements from the population ( $\sigma^2$ ) by the equation

$$\sigma_{\mu}^2 = \frac{\sigma^2}{n}$$

and so the number of tests needing to be carried out in order to obtain the required accuracy is such that

$$\frac{\sigma}{\sqrt{n}} < 0.0255\mu, \quad \text{or} \quad n > \frac{\sigma^2}{(0.0255\mu)^2} \quad (4.1)$$

$\mu$  and  $\sigma^2$  can be estimated by calculating  $\bar{x}$  (the measured average) and  $s^2$  (the measured variance) for a large sample. In this case (for the drives ASM)  $\mu$  and  $\sigma^2$  were estimated over 10000 tests to be equal to 6.44 and 41.19, thereby giving the required number of tests as  $(41.19/(0.0255 \times 6.44)^2) \approx 1500$  for the drives ASM. Experience with tests on other ASMs in the SE has shown that similar numbers of tests are also sufficient in those cases.

This result was tested by taking a population of 10000 individuals (i.e. 10000 measurements of the performance of the drives ASM), calculating the mean  $\mu$  and variance  $\sigma^2$  of the population and then calculating sample means  $\bar{x}$  for 5000 subsets of 1500 individuals. In 97.9% of the cases the values of  $\bar{x}$  were within  $\pm 5\%$  of the true mean  $\mu$ , which indicates that 1500 tests are sufficient to obtain a measurement of the average genetic fitness to the desired accuracy. The figure of 97.9% is somewhat greater than 95% because the individual measurements (genetic fitnesses) are forced into discrete values, whereas the measured averages ( $\bar{x}$ ) are not. This means that  $\sigma_{\mu}^2$  is actually somewhat less than  $(\sigma^2/n)$ .

The second characteristic to examine is the degree of certainty with which it can be stated that two ASMs truly differ in their performance in the SE, based on the average performance of each. If the two ASMs are tested  $n_1$  and  $n_2$  times respectively producing averages and variances of  $\bar{x}_1$ ,  $\bar{x}_2$ ,  $s_1^2$  and  $s_2^2$  then there is a significant difference between the two averages at the 95% level if

$$\left( \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right) > 1.96. \quad (4.2)$$

That is to say, if the inequality 4.2 holds then the observed difference between  $\bar{x}_1$  and  $\bar{x}_2$  is, with 95% probability, due to an underlying difference in performance rather than a random fluctuation between the results from two essentially similar ASMs. Similarly, there is a 99% certainty of a real difference if the LHS of 4.2 is  $> 2.576$ , and a 99.9% certainty if the LHS is  $> 3.291$ .

So, for instance, if  $s_1^2$  and  $s_2^2 = 14.0$ , and  $n_1$  and  $n_2 = 1500$  then there would have to be a difference between the two means of at least 0.268 for it to be significant at the 95% level. Similarly, a difference of at least 0.352 would be required for it to be significant at the 99% level, and a difference of at least 0.450 for it to be significant at the 99.9% level.

# Chapter 5

## Discussion

### 5.1 Assumptions

The major assumptions that were made during the construction of the SE, together with their implications, are now discussed:

1. **set of sub-problems chosen:** a particular set of 13 sub-problems was chosen to be representative of those faced by animals in general. While it is possible that the particular set chosen here is unbalanced in some way (in addition to the fact that there are no social sub-problems apart from mating), efforts were made to avoid this by including a large set of sub-problems and by considering the different ways in which sub-problems could vary (this is discussed in section 7.2).
2. **excluding social interactions:** the main simplification that has been made in the design of the SE is to keep social interactions to a minimum by including only one social sub-problem, that of mating. Although this is necessary in order to produce a manageable genetic fitness function, and in order to keep assumptions about the tested entity out of the testing procedure or 'testbed', it does make the SE somewhat less realistic of animal environments in general. While this drawback is noted, it should also be kept in mind that many animals (e.g. snakes, hedgehogs, tigers) do lead lives which are mostly solitary.

3. **discretising space and time:** in line with most other SEs, the environment was split up into a grid of squares. Although it would have been possible to represent the environment as a continuous space, this was decided against because of the complications that would have been added to the programming of the SE. Calculating the animal's perception and the interactions between the animal and predators, for example, would have become much more difficult. It was decided that the small amount of added realism was not worth the large amount of extra time it would have taken to include.

As with all other SEs, time in the SE is not treated as continuous but rather is split up into a series of timesteps. As with space it is possible, but much more complicated, to use a continuous representation of time, but it was decided against in this case.

4. **longer time-scale decisions:** the SE is only concerned with the moment-to-moment decisions an animal has to make. Changes in behaviour are often required due to maturation of the animal (e.g. sexual behaviour appearing after an animal becomes sexually mature) or occur in response to the time of year (e.g. migration). These longer-term 'decisions', usually mediated by hormones, will often have a very significant effect on the animal's behaviour, but are not considered here.

5. **adding edges:** a rather artificial assumption added to the SE was that of imposing fixed edges. Again, this is not strictly necessary in that the SE can be made big enough so that there is only a minute chance of the animal ever reaching the edge of it, but this would slow down the time it takes the program to run by a considerable amount, and would not add greatly to the realism. The graph in figure 4.7 shows that, with the drive-type ASM, the animal only reaches the edges of the SE infrequently, and so the addition of the artificial sub-problem 'avoid edges' does not interfere to any great extent with the action selection problem.

6. **splitting action selection off from other processes:** A more fundamental assumption made in this project is that it is possible to consider action selection

independently of perception, navigation and motor control. Could it be that the different processes are inter-linked so closely that the solution to any one process determines the form of the other processes? Does the way in which an animal perceives the environment (e.g. mainly by smell or vision, with varying ability to classify or generalise correctly, etc) crucially influence the form of the ASM? Or, similarly, does the way in which the animal executes its actions crucially influence the nature of the ASM?

While it is obviously the case that there will be some interdependency between different processes (at the interfaces if nowhere else), the essential nature of the problem of AS exists independently of the nature of the perceptual, navigational and motor control systems. However the animal perceives the SE, or executes its actions, it still needs to weigh up the needs of different sub-problems, assess their relevance to the perceived state of the environment, and to decide how to allocate its time. The form of the perceptual process will only affect the type of information coming into the ASM. The efficacy of perception might affect, for example, how often the animal should scan for predators, but it would not affect the need to do so at fairly regular intervals, and the need to fit scanning for predators in between other actions in a proper fashion. How the animal executes its actions (e.g. if it picks up food with its hands and then passes it into its mouth, or whether it just picks it up with its mouth directly) is relatively unimportant. The form of the navigational system may affect how far from the den the animal wants to stray, or how worthwhile it is for it to head to remembered food or water, but will not affect the overall nature of the problem of action selection. Different perceptual, navigational and motor control processes will lead to different instantiations of the *same type* of problem.

This argument for the independence of the nature of the action selection or time-allocation problem is supported to some extent by observations of animal behaviour. To quote from [Marler & Hamilton, 66]:

Perhaps the most remarkable characteristic of the principles of behavioural organisation is the variety of physiological mechanisms associated



with them. In breathing, feeding, and drinking, different mechanisms achieve similar ends. The temporal patterning of the feeding behaviour of flies and mammals is much the same, but the patterns are brought about by quite different physiological mechanisms.

If behaviour tends to be temporally arranged in similar ways in very dissimilar animals, ones which have very different perceptual, navigational and motor control processes, then it would seem likely that this is because the underlying nature of the problem is similar for these different animals. Just as the underlying constraints on vision have produced very similar, but separately evolved eyes in octopuses and mammals [Bruce & Green, 85], so the underlying constraints on action selection/time-allocation have produced mechanisms with similar effect in flies and mammals.

## 5.2 Using the Simulated Environment

Some considerations to keep in mind when testing ASMs with the SE are:

1. **checking for effects of particular models of perception, navigation and motor control:** although it was argued in the previous section that it is valid to look at AS as a separate process, different forms of the perceptual, navigational and motor control processes may lead to slightly different results for individual tests. If, for instance, ASM1 is better at dealing with the need to stay close to the den than ASM2, but worse in other aspects, then altering  $\sigma^2$  (the increase in variance of the animal's estimate of its position each time it moves) may alter the relative performances of ASM1 and ASM2. To check for effects of this sort it is a good idea to repeat tests in the SE with differing models of perception, navigation and motor control to see if the results are maintained.

Perception can be changed, for instance, by altering the way in which efficacy of perception changes with distance from the animal. Navigation can be changed,

for instance, by altering the value of  $\sigma^2$  or by altering the number of memories that can be remembered (the size of the map). Motor control can be changed, for instance, by altering the likelihood of actions being incorrectly executed.

2. **determining optimal parameter values:** any ASM will have its structure or underlying architecture specified with some degree of precision. However, all ASMs have parameter values which will be specific to the individual problem (e.g. parameters used to determine the importance variables in the drive ASM — each importance variable or drive strength will be a function of some internal, external and indeterminate stimuli and this function will contain parameters whose values are not specified, and which need to be tuned to the particular problem being faced).

The parameters can be set in one of two ways: either by hand (by using common sense and then observing the resulting performance and fine-tuning), or by using some search technique to optimise the parameter values with respect to the resulting genetic fitness. An attempt was made during this research to use genetic algorithms [e.g. Goldberg, 89] to optimise the parameters for each ASM. Despite the application of a lot of time and of extensive computing power this proved infeasible. Three reasons why it proved difficult to use genetic algorithms for this problem are:

- (a) **Low signal to noise ratio** : the variance of the performance measurements of the same ASM in the SE is fairly high in comparison to the differences in average performance for two different ASMs. This means that the rate of convergence towards a solution has to be slow to protect against the rejection of good solutions on the basis of a few 'unlucky' (i.e. below average) results.
- (b) **Number of parameters** : Many parameters have to be evaluated simultaneously (several hundred for some mechanisms).
- (c) **Interdependency of optimal parameter values** : genetic algorithms work best when individual parts can be optimised independently. This causes a problem for this application. The very nature of the action selection problem

is one of different needs competing for control or influence over the actions of the animal. For instance, in the drive mechanism an optimal solution will consist of the right balance of drives strengths of different systems in different situations. The optimal drive strength for 'get food' when the internal food deficit is very high depends partly on what drive strength occurs for 'clean' when uncleanliness is very high, on what drive strength occurs for 'avoid predators' when a predator is near, etc. The optimal values of each drive strength in each stimulus situation are all interdependent. If the drive strengths are doubled for ten of the sub-problems (in all stimulus situations), then the values of the remaining drive strengths will also need to be doubled in all situations to become optimal again. It is the *relative* rather than the *absolute* values of the drives strengths that are important.

The difficulties with using an automatic search method meant that parameter values had to be set by hand and fine-tuned using trial and error, although this was less preferable.

3. **number of tests to be carried out:** as argued in section 4.3, each test should involve at least 1500 trials in the SE in order to produce an average genetic fitness which is 95% certain to be accurate to within  $\pm 5\%$ .

### 5.3 The Fallibility of Results from the Simulated Environment

One issue arising out of the previous two sections is the 'level of proof' that should be attached to results from the SE. Several factors mean that the results from the SE should be treated somewhat sceptically: (i) the choice of the sub-problems in the SE is biased away from social interactions, (ii) the authenticity of the models of perception, navigation and motor control (as well as of the sub-problems themselves) cannot be guaranteed, even though they have been carefully constructed, (iii) the parameter values in each action selection mechanism have to be set by hand and so cannot be guaranteed to be optimal.

Because of these three factors, results in the SE by themselves should not be considered as *proof* that an ASM would not be able to select actions appropriately in the real world, or that one ASM is better than another. Instead, the SE should be used as a tool with which to look for shortcomings with mechanisms and to get an approximate idea as to the relative performances of different mechanisms. Once a deficiency in the performance of a mechanism is highlighted by the SE then the underlying reasons for that deficit can be investigated, and it can be shown to be either due to an inherent deficit in the ability of the mechanism to select actions appropriately or else due to a peculiarity of the SE (or a peculiarity of the dynamics of the interactions between the ASM and the SE). An analogy can be made with the old practice of keeping a canary down a coal mine. If the canary became unconscious then this did not infallibly prove the existence of dangerous gasses in the mine, but gave a strong indication of them. Even though canaries were rather unreliable detectors, this did not rule out their usefulness before more accurate and infallible measuring devices were available.

In the same way, the SE can be used to pin-point potential inadequacies with ASMs (and it will be shown in the second half of the thesis that it is able to do this), but it cannot by itself *prove* the existence of those inadequacies. When two mechanisms show different results in the SE then, similarly, that by itself cannot be used as proof that one is better or worse than the other. Only if the difference in performances can be traced to a fundamental difference between the two mechanisms (one that it can be shown is relevant to action selection in general, rather than just in the SE) is it proven that one mechanism is better than another at selecting actions.

## 5.4 Terminology

The terminology described in the Prologue, and used throughout this thesis, is partially taken from that used by Baerends [1976]. The terms system, sub-system and action are different from those proposed by McFarland & Sibly [1975].

McFarland proposes three terms, all of which are for entities on the 'solution side' (see figure 0.4): (i) A *characteristic* is an atomic movement such as pecking or swallowing.

(ii) An *action* is a regularly occurring, distinguishable, sequence of characteristics such as 'eating grain', which in the hen might consist of pecking followed by mandibulation (moving grain from beak to back of throat) followed by swallowing. There is a further stipulation that actions must be mutually exclusive, i.e. no two actions can be assigned such that they can occur simultaneously. (iii) An *activity* is a grouping of actions which is only valid if it is composed of actions that belong solely to that activity. Therefore activities are also mutually exclusive because they are composed of non-overlapping sets of mutually exclusive actions. Activities are not necessarily functionally-related sets of actions (i.e. they will not necessarily be conveniently labelled groups such as 'feeding').

McFarland & Sibly's terminology has not been used here partly because of its complexity. This terminology was intended mainly for the use of ethologists classifying the *observed behaviour* of animals. Baerends' grouping together of functionally-related entities into systems and sub-systems is more suitable for referring to ASMs and their components.

'Functionally-related' is not a specific enough term for deciding what should and should not be included in the same system. The term has been interpreted here to mean (i) the entities in a system are relevant to the same function, and (ii) the entities in a system share a common causal factor (i.e. stimulus). This means, for instance, that vigilance and predator avoidance are two separate systems. The first system has two common causal factors, affecting all entities in the system, which are the time since the animal last scanned its environment for predators and the time since a predator was last perceived. The second system has a different common causal factor related to the degree of proximity of a predator. Although the two systems can be thought of as subserving a common higher-level goal (not getting eaten by predators), they do not share causal stimuli and so are classified as separate systems here. [Hinde, 66] contains a good discussion of the distinction between functional classifications and causal factors.

## 5.5 Theoretically Optimal Actions

In simple, deterministic simulated environments it is possible to calculate the optimal action for the animal at each moment in time. Even when the transitions between states (dependent on selected actions) are only specified probabilistically, then the optimal action at each point in time can still be determined, if the animal has a fixed life-span. Some techniques that are applicable to this problem are Markov processes, dynamic programming, finite state automata and Pontryagin's Maximum Principle [e.g. McFarland & Houston, 81].

If the state of the environment (both the external environment and the animal's internal variables) is currently  $S_i$  and the animal performs an action  $A_j$  then let the probability of the state of the environment changing to  $S_k$  be

$$P_{ijk} = P((S_i, A_j) \Rightarrow S_k)$$

Also, let the future expected genetic fitness of each state  $S_i$  be  $R_i$ , the optimal action be calculated for time  $t_0$ , and the time the animal dies be  $t_D$  (a fixed length lifespan is assumed). Given that the *actual* genetic fitness is known for each state at time  $t_D$ , it is originally possible to calculate the optimal action only for each state at  $t_{D-1}$ . This will be the action  $A_j$  for each  $S_i$  which results in the greatest value of the expression

$$\sum_{k=1}^{k=N_s} (P_{ijk} R_k)$$

where  $N_s$  is the number of different states (the size of the state space) and  $R_k$  is the actual genetic fitness of each state  $S_k$  at time  $t_D$ . It is assumed that the animal will always choose the optimal action (i.e. will follow what is known as an optimal policy). If this is true then the maximum value of  $\sum_{k=1}^{k=N_s} (P_{ijk} R_k)$  becomes the *future expected genetic fitness* ( $R_i$ ) for each state  $S_i$  at time  $t_{D-1}$ .



As an example, imagine a state  $S_A$  and two possible actions in that state  $A_X$  and  $A_Y$ . Performing  $A_X$  in  $S_A$  has a 0.7 probability of changing the state to  $S_B$ , which has a genetic fitness of 12.0, and a 0.3 probability of changing the state to  $S_C$ , which has a genetic fitness of 3.0. Performing  $A_Y$  in  $S_A$  has a 0.4 probability of changing the state to  $S_B$ , and a 0.6 probability of changing the state to  $S_C$ . In this example the future expected genetic fitness of the two actions would therefore be  $((0.7 \times 12.0) + (0.3 \times 3.0)) = 9.3$  for  $A_X$  and  $((0.4 \times 12.0) + (0.6 \times 3.0)) = 6.6$  for  $A_Y$ , and so action  $A_X$  is optimal for state  $S_A$ . Assuming the animal follows an optimal policy then  $R_A$ , the future expected genetic fitness of state  $S_A$  at that time, has a value of 9.3.

In the same way that the future expected genetic fitness of each state at  $t_{D-1}$  can be calculated from those at  $t_D$ , so it is possible to continue stepping back through time calculating the optimal actions for each state at times  $t_{D-2}, t_{D-3}, \dots$  (choosing the action which maximises  $\sum_{k=1}^{k=N_s} (P_{ijk} R_k)$  at each timestep), until  $t_0$  is reached, and the optimal action for the current state at the current moment in time can be calculated.

It would be useful if some technique along the lines of that just described could be used to calculate optimal actions in the SE. These optimal choices could then be compared against the actual actions chosen by ASMs. However, there are several properties of complex and realistic environments which make it difficult or impossible to calculate the optimal actions for them:

1. the size of the state space is enormous. Even if those stimuli which are continuous variables are discretised, then the state space is completely unmanageable. For instance, the size of the state space for this SE is of the order of  $10^{100}$ , even assuming that all of the real-valued stimuli are discretised fairly coarsely. For a more realistic and more detailed simulation the state space would be many orders of magnitude larger.
2. the animal's lifespan is not of a fixed length.
3. the transition probabilities  $P_{ijk}$  cannot be accurately calculated.

Some work has been done on looking at how the effects of an unmanageably large state space can be reduced (for example, [Moore, 90]) and on how optimal policies can be calculated for animals with lifespans that are not fixed in length, but even so these difficulties, especially the first and third, mean that it is not feasible to use a purely mathematical approach to calculate optimal actions for the SE.



## Chapter 6

# Conclusions to Part I

This first part of this thesis has described currently existing simulated environments and explained why a simulated environment is a sensible and valid approach to studying action selection. A new simulated environment, more realistic and complex than current alternatives, has been developed in a principled manner with the underlying theory and motivating principles explained. The thorough and extensive testing procedures used to validate the simulated environment were described. Finally, practical issues concerning how to use the simulated environment to test action selection mechanisms were discussed, and the effects of some of the assumptions made in its design were considered. In particular, it was argued that the results from the simulated environment are fallible and should only be used as *indicators* of shortcomings with a mechanism or of a real difference in performance between two mechanisms.

The simulated environment presented here contains error-prone models of perception, navigation and motor control; 14 different sub-problems; 14 different types of feature; 6 internal, 6 indeterminate and many external stimuli; and 35 different low-level actions to select between.

The simulated environment presents a demanding and rigorous test of action selection. It does have limitations, which have been discussed at length, but these do not

preclude its usefulness as a tool for investigating action selection mechanisms. Its usefulness will be demonstrated in the second part of the thesis in which the inadequacies of several action selection mechanisms are uncovered with its help. There are two reasons that make this simulated environment a valuable and probing instrument for the examination of action selection mechanisms: (i) great care was taken to limit the number of explicit and implicit assumptions that were made, and the implications of those assumptions that had to be made were carefully considered. The simulated environment is a principled and painstaking embodiment of the action selection problem for animals, rather than an overly-simplified abstraction of it. (ii) the simulated environment provides a means for verification and testing in a field where it has not existed before. Previously, the only means of analysing the performance of models or theories of action selection (such as Lorenz's "psycho-hydraulic model") has involved 'thought experiments' or limited mathematical analyses. This simulated environment allows a more challenging, thorough and penetrating means of validation.

## 6.1 Further Work

Some ways in which the work in this part of the thesis could be extended are as follows:

1. The SE could be rewritten to work with a continuous representation of space rather than the current grid-based system.
2. The SE could be made more realistic in other ways. For example, more sub-problems could be included and more realistic models of the low-level mechanics of perception, navigation and motor control could be added.
3. The SE could be adapted to model the specific action selection problems for individual types of animal, in addition to the generalised model here. For instance, a computer simulation of the action selection problem for the herring gull could be implemented. This could make use of the fairly extensive information describing the contingencies of the herring gull's environment that is available in sources such as [Baerends & Drent, 76], [Baerends & Drent, 82] and [Sibly & McCleery, 85].

4. In the sense that actions are defined in this thesis (they are not mutually exclusive *by definition*, in contrast with the actions of McFarland & Sibly [1975]), the assumption of the behavioural final common path (that all actions are mutually exclusive) is not always true for all animals (e.g. some animals can eat and walk simultaneously). For most animals many actions are incompatible but some can be executed simultaneously. This could be reflected in the SE.

## **Part II**

# **Evaluation of Action Selection Mechanisms**

## Chapter 7

# Introduction

Part I of this thesis has dealt with the simulated environment. This second part now looks at different action selection mechanisms (ASMs). Several existing ASMs are described, some of which were implemented and tested in the SE. For those that were implemented and tested, their performance in the SE is described, and the underlying reasons behind any gross inadequacies in their performance are discussed. Following on from the experiences of shortcomings with different mechanisms, a list of essential requirements for any successful mechanism is presented. Finally, an ASM is described which has been designed in the light of the failures and successes of other mechanisms, and which is shown to perform better in the SE than all the others that were tested. The theoretical underpinnings of this mechanism are explained in depth.

### 7.1 Scope of Testing Procedure

The testing procedure in the SE is designed to measure the ability of an ASM to produce optimal or near-optimal selections of actions. The mechanisms are evaluated in terms of their computational properties, rather than their abilities to reproduce animal behaviour phenomena. The mechanisms put forward by ethologists are designed partially to account for observable phenomena in animal behaviour such as time-sharing

(the appearance of gaps in a type of behaviour which should be dominant over a long period of time), displacement activity (the occurrence of seemingly irrelevant behaviours when two relevant types of behaviour such as aggression and escape are vying for expression), and vacuum activity (the occurrence of actions which have not been executed for a long time, even though they appear irrelevant to the current situation). Testing these mechanisms for computational optimality is useful, but it should be remembered that this is not the only criterion by which they should be judged.

The only learning to be considered here is that of learning the location of features in the environment. The ASM receives navigational inputs which will change over time as the animal encounters and remembers different features, but these inputs are not under control of the ASM and, apart from in response to these navigational inputs, none of the mechanisms considered change their selected actions in the light of experience. The decision to exclude learning from the ASMs was made in order to simplify the problem being studied. It should be noted that no *tabula rasa* learning system (such as Sutton & Barto's Temporal Difference Learning system [Sutton, 88] or Watkins' Q-Learning system [Watkins, 89] or Booker's classifier systems [Booker, 88]) would perform well in the SE. Experiments with approaching predators or hazards do not usually allow for learning through experience because the animal will die on the first experiment. Although trial and error and learning from experience obviously do play a part in real-life action selection in more advanced animals, its importance in general has perhaps been over-emphasised. In situations where the 'right' solution can be pre-programmed or hard-wired, then it is obviously inefficient to require it to be learnt through trial and error. This is true even for less critical aspects of behaviour such as deciding what actions will increase the animal's cleanliness.

Another limitation on the testing procedure used here is that there is no consideration of decisions over much longer time-scales (e.g. those usually mediated by hormones in animals). In the SE there are no seasons or maturational changes in the animal, and so there is no need for the animal to make long-term decisions such as when

to hibernate, when to enter reproductive phase or when to change from behaviour patterns appropriate for a young animal to those appropriate for an adult.

A final point is that the amount of computation required for each decision is not penalised here. In nature, of course, evolution will act to minimise both the time taken to compute action selection decisions and the size of the nervous system required to compute them.

## 7.2 A Taxonomy for Sub-Problems

As mentioned in Part I, the main problem for an animal (reproduction and hence survival) leads to many subsidiary problems, such as getting enough food and avoiding predators, each of which differ in the nature of the demands they place on the animal's time. Before proceeding with the rest of this part of the thesis it is helpful to try and characterise the different sub-problems and to develop a *classificatory scheme* with which the sub-problems can be described. The purpose of doing this is two-fold: firstly the descriptive terms will be useful as of themselves, and secondly the process is useful in order to give some idea of how much

variation there can be between different sub-problems.

Figure 7.1 shows how the 'urgencies', or 'potential benefits', for six different sub-problems might vary over time. The black circles denote occasions on which the selected action is chosen because of the needs of the relevant sub-problem. These six sub-problems are now described:

1. Trying to obtain enough food is the most commonly covered sub-problem in the ethological literature. The likelihood of the animal trying to obtain food should depend on both external stimuli (whether it sees food in the vicinity) and on internal stimuli (whether it is short of food). It is also a recurring, continual problem. An animal will need to eat a certain average amount each day. The

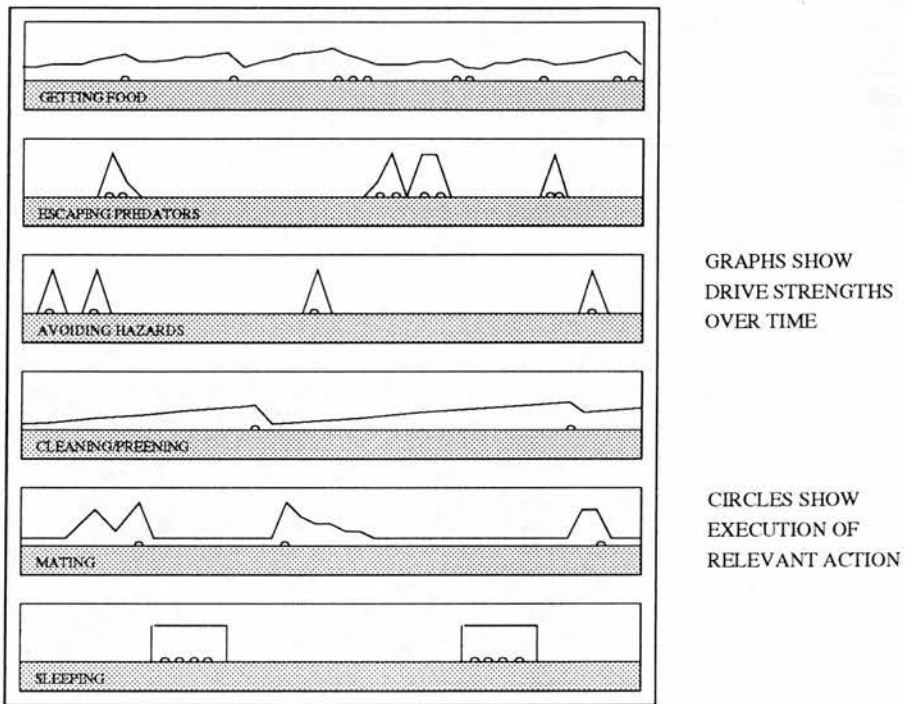


Figure 7.1: 'Longitudinal Profiles' for the six example sub-problems. The graphs show the drive strengths of the sub-problems, the highest of which at each moment in time wins the competition. The circles denote instances when the sub-problem wins the competition and therefore determines the animal's action (after [Maes, 91a]).

need to obtain food will in general place a low-urgency demand on the animal's time, unless the animal is particularly short of food, since the animal is not likely to die suddenly or become injured if it does something else. Finally, food intake is a homeostatic problem.

2. A second common sub-problem is the need to escape predators. This is usually urgent and overriding, in that if the animal does not attend to it immediately



then the consequences could be fatal. It is also highly dependent on external stimuli but not at all on internal stimuli. The priority which the animal should assign to trying to escape from a predator is highly dependent on whether the animal senses any predators, and if so how close they are. This sub-problem is a non-periodic, intermittent sort in that there is no pattern to how often the animal will need to attend to it. An animal may need to escape from a predator twice in a short interval of time and then not need to do so again for a long period thereafter. There is no homeostatic aspect to this sub-problem.

3. A third sub-problem is that of avoiding hazards in the environment – places where an animal will endanger itself if it goes there (e.g. cliffs, streams). When an animal is near to one of these it is important that it does not move towards it. The demand on the animal's actions is *proscriptive* ('ruling-out'), rather than *prescriptive* ('specifying'), as is the case for other sub-problems. A proscriptive sub-problem specifies that certain actions should *not* be chosen (e.g. do *not* move towards a hazard) rather than that they should be chosen (e.g. eat food). This sub-problem will be urgent, but does not preclude the choice of another action at the same time. There is no homeostatic or periodic or continual aspect to it.
4. Another common sub-problem is cleaning, preening or grooming. Most animals need to spend some time every so often to remove dirt/parasites from their fur, clean and oil their feathers, or whatever. This will not be an urgent activity, since it will not be crucial to the animal to pay attention to it at any particular moment in time. It will tend to occur most frequently at moments when no other activity is urgently required. It is continual in that the need for it will recur frequently. It is dependent on internal stimuli but not on external ones.
5. A fifth sub-problem is that of mating. External stimuli are important in that the animal should attach more priority to this sub-problem when a potential mate is perceived (assuming an animal which makes occasional matings with different mates and which forms no long-term partnerships). It is sometimes periodic and related to internal stimuli (e.g. menstrual cycles) and sometimes

not. This sub-problem will be prescriptive, non-continual and non-homeostatic. It will probably be fairly urgent but the level of urgency in relation to other sub-problems will depend on factors such as how often opportunities for mating arise and how much longer the animal can expect to live.

6. A final sub-problem is that of the animal needing to return to its den and sleep there at night. This is obviously periodic. It will have increasing urgency as nightfall approaches. It is non-homeostatic and prescriptive.

Six different sub-problems that compete for a 'share' of the animal's time have now been described. Some of the descriptions are probably not completely accurate but that is not important here. The important point is that competing demands on an animal's time (from different sub-problems) vary in their nature, and that the differences between them need to be understood. A vocabulary for describing sub-problems is now proposed:

1. **Homeostatic v. non-homeostatic** – a homeostatic sub-problem contains an internal variable which has a desired optimal value, or at least a desired range of values. The behaviour of the animal will always act so as to return the value of the variable towards the optimal value or range of values [Toates, 80, Mrosovsky, 90].
2. **External stimulus dependent v. external stimulus independent** – the urgency with which certain sub-problems should 'demand' the animal's attention is dependent on the appearance of certain external cues (e.g. getting water on the stimulus of a water source, escaping predators on the appearance of a predator). Other sub-problems are independent of external cues (e.g. cleaning).
3. **Internal stimulus dependent v. internal stimulus independent** – as for above, except that the important factor is an internal cue (e.g. body heat is too high, not enough food in the animal's stomach).

4. **Periodic v. non-periodic** – some sub-problems such as sleeping at night are highly periodic, with the desirability of paying attention to them rising and falling with a regular rhythm.
5. **Continual v. occasional** – some sub-problems need to be attended to frequently and the need for them keeps recurring (e.g. cleaning, getting food/water). They will need to be undertaken at least several times each day. They are often internal stimulus dependent. Others occur more occasionally and are usually external stimulus dependent (e.g. escaping predators, mating).
6. **Degree of urgency** – some external stimulus dependent sub-problems arise only occasionally but are extremely urgent and over-ridingly important when they do occur (e.g. escaping predators, avoiding hazards). There will be significant consequences for the animal (in terms of future expected genetic fitness) if the sub-problem is not allowed to influence the action the animal selects. Some sub-problems generally have a fairly low urgency (e.g. cleaning), and tend to take over only when none of the more urgent sub-problems are relevant. This distinction is somewhat similar to that between squashable and resilient behaviours [Houston & McFarland, 80].
7. **Prescriptive v. proscriptive** – most sub-problems require a certain set of actions to be carried out (e.g. find food, approach it then eat it), whereas others (e.g. avoid hazards) only require that certain actions should *not* be carried out.

This list is almost certainly incomplete in that there are other ways in which sub-problems can vary, but hopefully the most consequential differences are contained here. However complete or incomplete the list above, there is an important underlying point: sub-problems differ in many ways, and simple mechanisms which make simple assumptions, for instance that all sub-problems are homeostatic, will not suffice.

### 7.3 Other Terminology

Following on from the terms just developed for describing sub-problems, some other terms are now explained:

1. **consummatory** – a consummatory action is one which directly achieves something (like eating food, or copulating). In general the execution of a consummatory action will reduce an internal deficit or will directly improve something about the animal's environment.
2. **appetitive** – an appetitive action is a preparatory one (like approaching some fruit type food, or making a courtship display at a mate). There is no intrinsic value in an appetitive action as of itself. It is only of value in that it tends to bring about a situation in which a consummatory action can be performed.
3. **connectionist** – this term implies the use of nodes and connections, although it does not necessarily imply that the nodes or connections are particularly similar to biological neurons or synapses.
4. **hierarchical** – the definition of a *hierarchy* used here is taken from [Dawkins, 1976]. For two nodes *A* and *B* then *A* 'is boss of' *B* if *A* has a direct causal influence on *B*. *Direct* is meant to imply that *A* is immediately above *B* in the hierarchy (i.e. its influence on *B* is not mediated through another node), and *causal influence* is meant to imply that the state of node *B* is to some extent dependent on the state of node *A*. For a set of nodes *A*, *B*, *C*, etc, *A* is said to be *superior* to *B* if (1) *A* is boss of *B*, or if (2) *A* is boss of a node which is superior to *B*. Using this recursive definition of the term superior it is now possible to define a hierarchy as a set of nodes *A*, *B*, *C*, etc. which satisfy:
  - (i) there is no node in the set which is superior to itself, and
  - (ii) there is one element in the set which is superior to all others in the set.

Figure 7.2 shows some example hierarchies. The hierarchical mechanisms discussed here are all composed of sets of overlapping hierarchies (one for each system) as in figure 7.2(e). They are sets of overlapping hierarchies because there are several 'top' nodes. For simplicity though, the different mechanisms will be referred to as hierarchies, although strictly speaking they are each sets of overlapping hierarchies.

5. **reactive** – a reactive mechanism is one which responds promptly to changes in the state of the environment and one in which the decision as to which action to select is determined mostly by the *current* state of the internal and external environment.
6. **deliberative** – a deliberative mechanism, on the other hand, is one in which the decision as to which action to select is based to a larger extent on *past* events, calculations or plans.

## 7.4 Plan of Part II

The rest of this part of the thesis is organised as follows: Chapter 8 describes many different ASMs, together with other relevant literature. Chapter 9 describes the results of trying to implement and test several of the ASMs in the SE, and examines the root causes of any gross inadequacies in performance. Chapter 10 looks in greater depth at some of the theoretical issues related to inadequacies in performance and also contains a list of essential requirements for an ASM, developed from analysis of the faults in the ASMs tested. Chapter 11 presents some suggestions as to how action selection might best be computed. The proposed mechanism is a development of one of the mechanisms described in chapter 8. Finally, chapter 12 contains some conclusions on this part of the thesis.

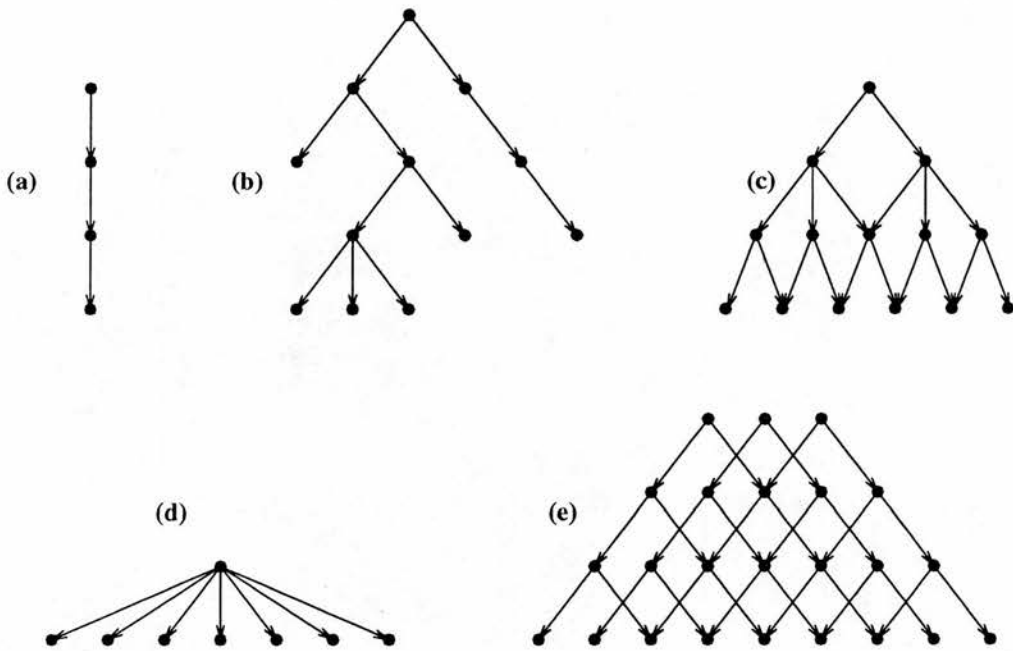


Figure 7.2: Example hierarchies. (a), (b) and (d) are non-overlapping hierarchies, (c) is an overlapping hierarchy because some nodes have two bosses, and (e) is a set of overlapping hierarchies (after [Dawkins, 76]).

---

## Chapter 8

# Mechanisms for Action Selection

This chapter describes the most important action selection mechanisms, as well as other work relevant to action selection. Some of the mechanisms described here were tested in the SE, and these are described in greater detail. The decision as to which ASMs to implement and test was made partly according to how completely and unambiguously they have been specified. Another consideration was the desire to test a representative sample of the different ASMs. This chapter contains a brief description of work from related fields, such as planning, and an explanation of why work from these fields is not considered in greater depth here. The literature reviewed comes mainly from the two disciplines of ethology and robotics, but also in part from other fields such as AI and psychology. Some of the mechanisms are connectionist, others symbolic. Some mechanisms are hierarchical, others not. Some of them address the whole of the problem of action selection while others deal only with parts of the problem. The mechanisms are presented in more or less chronological order of publication.

When looking at the descriptions of mechanisms in this section it should be kept in mind that some of the deficits that mechanisms are criticised for in chapters 9 and 10 are: (i) an inability to properly integrate goal or motivation type information (what the animal needs to do) with environmental or cue type information (what the

animal can usefully do in the current situation), (ii) an inability to choose appropriately between consummatory and appetitive alternatives, whether in the same system or in different systems, and (iii) an inability to deal successfully with all different types of sub-problem, as just discussed in section 7.2.

All of the ASMs use the assumption of the **behavioural final common path** [McFarland & Sibly, 75], that all *actions* are incompatible in terms of their demands on the effectors of the animal (the animal is not able to move its body so as to achieve any two of the actions simultaneously). This means that at the level of actions, an ASM has to make a decision as to which one action is most appropriate for the situation it finds itself in.

## 8.1 Drives

The concept of drives is associated most with Hull [e.g. 1943], but they have been advocated by many others in many different forms. Drives have also been discussed and criticised by many including Bolles [1975], Hinde [1959, 1960, 1970] and McFarland [McFarland & Sibly, 1972; McFarland, 1985], who also refers to action selection with a drives-like system as 'simple motivational competition'.

The term *drive* is used most commonly to denote the idea of an intervening variable. Many stimuli will have an effect on how likely an animal is to perform a particular type of behaviour. The term 'a particular type of behaviour' in this instance implies any action from a system. So, for instance, performing feeding behaviour implies choosing an action to explore for food, approach perceived or remembered food, or eat. Stimuli relevant to the actions in the system can each act separately to bring about the execution of the different actions, or they can act by way of an intervening variable or centre (see figure 8.1). A drive is equivalent to the activation of such an intervening variable and as the value of the drive increases, it will become more likely that one of the actions of the system will be chosen. The drive for each system is a measure of the total motivation, combined from all stimuli, for that type of behaviour.



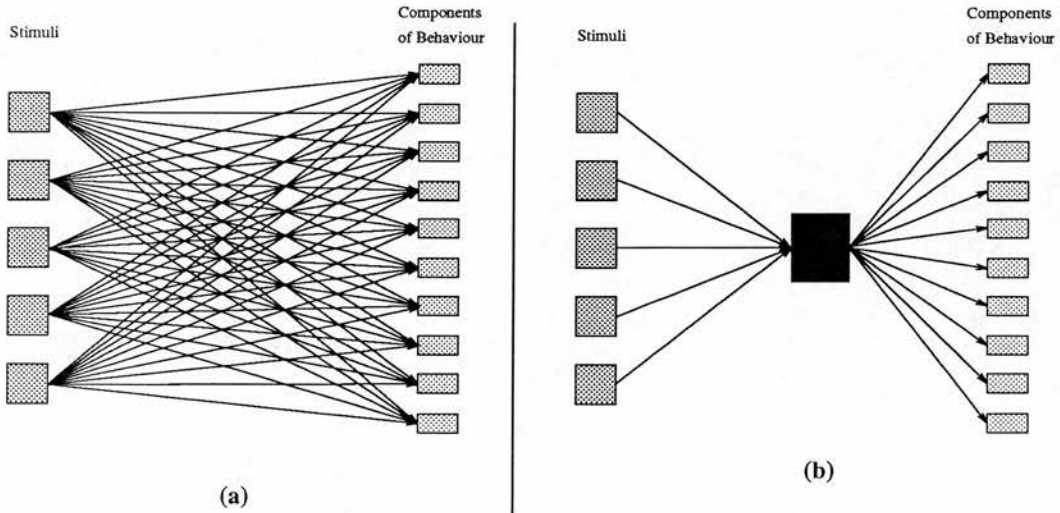


Figure 8.1: **(a)** many stimuli each passing activation to many components of a behaviour. **(b)** an intervening variable or drive (the black box).

There has also been much use by ethologists of the term 'general drive', i.e. arousal which is not specific to any one system but which makes the animal more likely to engage in energetic behaviour of any sort. In this thesis the term 'drive' is restricted to the more common meaning in the previous paragraph.

Some formulations of drives theory involve learning. Hull proposed a central equation, for each system, of the form

$$drive\_strength = habit\_strength \times stimulus\_strength$$

where the habit strength is the result of learning from previous experience and the stimulus strength is some unspecified combination of internal and external (and indeterminate) stimuli. Removing the influence of learning gives

$$drive\_strength = stimulus\_strength = f(internal, external\ and\ indeterminate\ stimuli)$$

where  $f$  is some unspecified function. Many formulations have also implicitly assumed that  $f$  is a function of internal stimuli only (e.g. the drive for feeding would be a function of the internal deficit only).

The only remaining unmentioned component of the drive concept is the central assumption that the most important system is calculated by comparing respective drive strengths and choosing the system with the highest strength. The most appropriate *action* for that *system* alone is then chosen in some unspecified manner. It is implicitly assumed that the hard part of the problem is choosing which system to concentrate on, and that choosing a suitable action thereafter is not difficult.

Figure 8.2 shows a list of drive strengths, one for each system, calculated from environmental, internal and indeterminate stimuli. In this case feeding behaviour would be chosen because it has the highest value drive.

It is apparent that while the concept of drives is very relevant to the AS problem, it leaves crucial questions unanswered. For instance, how should one combine different stimuli to calculate the value of the drive variable and how should one choose the most appropriate action once a system has been selected?

As a final note, one obvious possible shortcoming of drives has been pointed out by many authors. For systems such as getting food and getting water it is necessary to calculate the drive strength using both internal (e.g. food or water deficit) stimuli and external (e.g. food or water perception) stimuli. Otherwise a phenomenon known as **dithering** (see figure 8.3) can occur. Imagine an animal with equal food and water deficits and situated between a food and a water source. If perceptual stimuli (which increase with proximity) are not allowed to increase the relevant drive strengths then the animal might visit the food source and decrease its food deficit slightly. Now the drive for water dominates over that for food and so the animal moves back to the water source, drinks a little water and finds that its food deficit is now higher again. An oscillation occurs between the food and the water with the animal spending far more time moving between them than either eating or drinking.

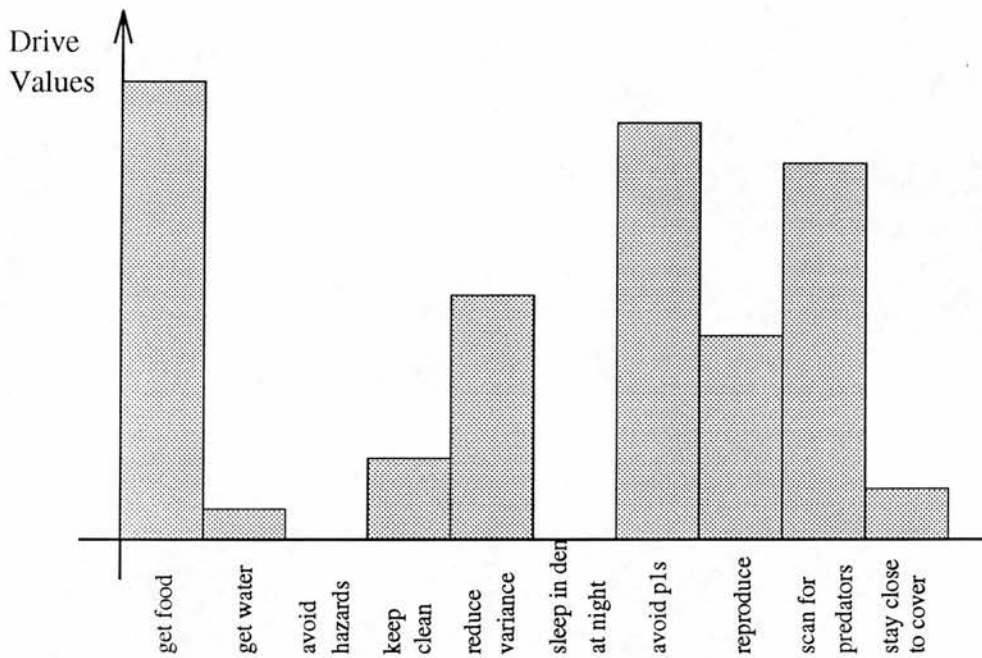


Figure 8.2: Example drive strengths for a drives ASM, with 'get food' system winning.

By including external stimuli in the calculation of drive strengths, the animal will become more attracted to food or water sources as it gets closer to them, and will remain at each source for longer (beyond the time when the respective deficit is no longer the largest). In this way it will spend more time consuming and less time travelling between the different food and water sources.

The drives model, or an instantiation of it at least, is one of the ASMs that were implemented and tested in the SE. A more precise description of how it was implemented and tailored to the particular problem posed by the SE is given in section 9.3.1.

## 8.2 Tinbergen

Tinbergen's mechanism [Tinbergen, 50 and 51] is more precisely described than the drive model but is still rather ambiguous in places. It was intended only as a 'working

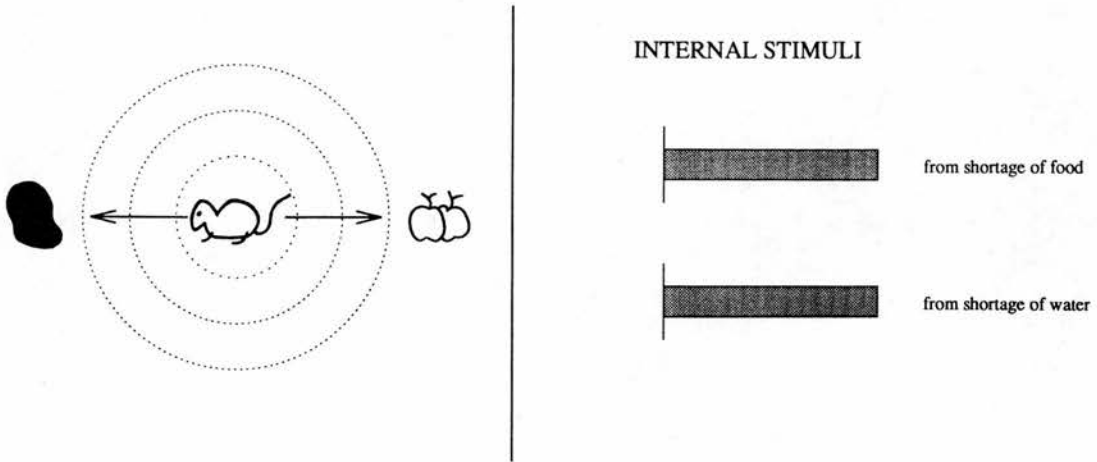


Figure 8.3: Dithering – with equal food and water deficits, the animal may oscillate rapidly between small bouts of drinking at the water source and small bouts of feeding at the food source.

hypothesis’ and as such is not a very detailed or exact specification. The mechanism is hierarchical. At the highest level are very general nodes such as ‘reproduce’ and ‘get food’. As one proceeds ‘down’ the tree the nodes in turn represent successively lower-level and less abstract concepts. At the lowest level are consummatory actions such as coition and eating. This mechanism addresses the whole of the action selection problem as well as longer term decisions such as when to enter ‘reproductive phase’.

Figure 8.4 shows a part of the hierarchy for one system. The idea is that activation energy comes in at the top node of each system from motivations, which usually correspond to internal stimuli, but sometimes also to external stimuli. This activation is then allowed to pass on to lower-level nodes in turn as long as the node sending the activation is not ‘blocked’. A node is ‘blocked’ when its *innate releasing mechanism* is not active. The innate releasing mechanisms are controlled by external stimuli and act to unblock only those nodes which are relevant to the current situation.

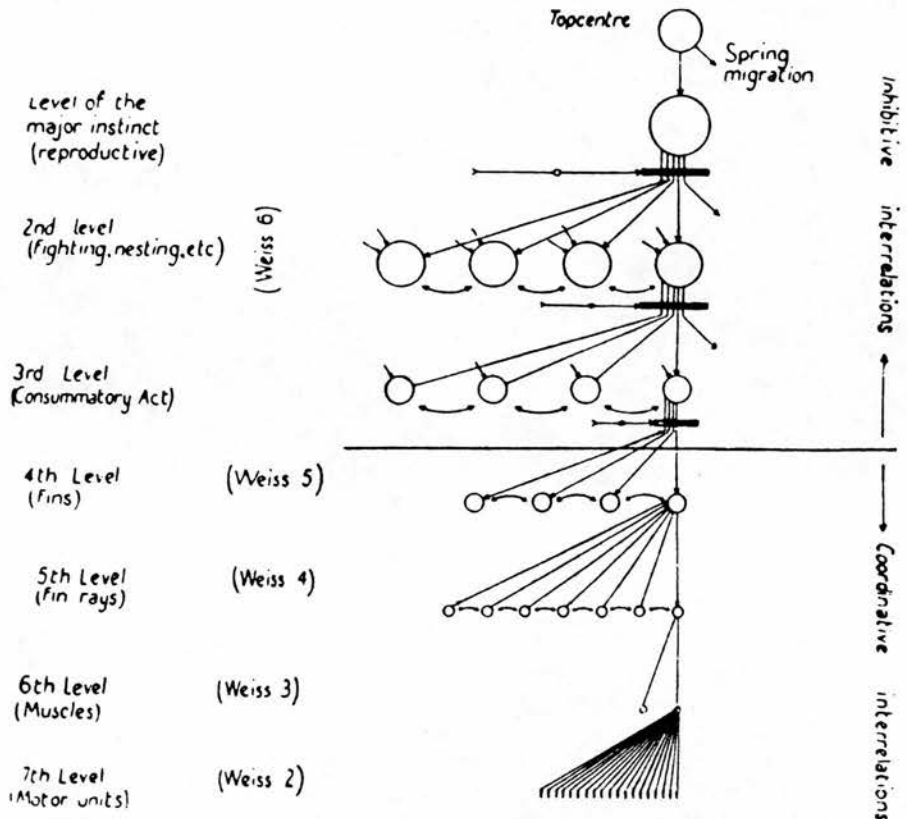


Figure 8.4: Tinbergen's 'Hierarchy of Instincts' (the bottom half shows motor control and can be ignored). Circles = nodes or centres, downwards arrows = causal factors, horizontal arrows = innate releasing mechanisms, and two-way arrows between centres on the same level = mutual suppression. The bottom half of the diagram (Weiss 5 to Weiss 2) is concerned with motor control and is not relevant here. Taken from [Tinbergen, 51].

For instance, the innate releasing mechanism for the 'fight' node of a male stickleback might be the external stimulus of another male invading its territory. If this stimulus exceeds a critical value then the innate releasing mechanism will fire and the activation in the node can pass down to lower level nodes. The stimuli impinging on each innate releasing mechanism are combined according to a "rule of heterogeneous summation" (i.e. a weighted sum is calculated).

In addition to needing to be unblocked, each node's activation must exceed a separate

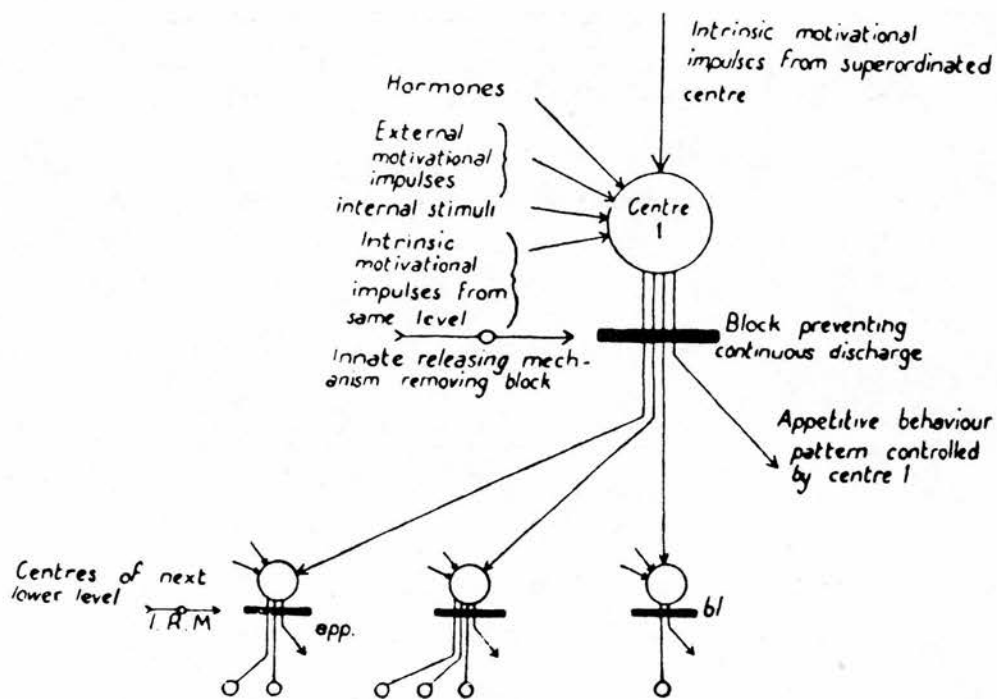


Figure 8.5: Tinbergen's 'tentative representation' of a node or centre from an intermediate level in the hierarchy. Taken from [Tinbergen, 51].

threshold in order for its activation to filter on down to lower levels. As shown in figure 8.5, a node receives activation from many sources such as motivational impulses from a higher level and external stimuli. There is also intra-layer inhibition which can decrease a node's activation. How exactly this should work is unspecified, but the inhibitory links should be arranged so that "as a rule, the strong activation of instinctive behaviour of one kind prevents the functioning of another pattern" [Tinbergen, 51, page 111]. It is also stated that "the threshold for the activation of the highest [furthest from consummatory] appetitive behaviour must be higher than that for the activation of the low type of appetitive behaviour" [Tinbergen, 50, page 310]. It is unclear exactly what is intended except in that there should, usually at least, be a *winner-take-all* competitive process operating at each level in the hierarchy.

In the description of the mechanism there is an implicit analogy with fluid filtering down through some contraption which possesses various removable gates. When

activation energy is released from a node by an innate releasing mechanism, then that node's gate is lifted and the energy is assumed to 'flow down' to the lower levels and be shared amongst them. When a consummatory action is able to be performed then it is assumed to 'drain away' the fluid from that part of the hierarchy, and in particular from associated appetitive actions which are now no longer relevant.

The levels in Tinbergen's hierarchy do not correspond exactly to gradations between system-level nodes and action-level nodes (see figure 0.3 for explanations of terms). Although there is some indication of this it is also stated (e.g. figure 8.4) that consummatory actions are invoked by the lowest level nodes whereas appetitive actions are invoked by successively higher level nodes, as the appetitive actions become more and more remote from the consummatory act. As a whole the descriptions of this mechanism are vague and informal as one would expect of a "working hypothesis". A more formal and mathematical approach would have been helpful in terms of presenting a less ambiguous definition of the mechanism. Tinbergen's model was not implemented and tested in the SE, mainly because of this ambiguity in its specification. Due to the importance of hierarchical models of this type in the literature, a more generic ASM, a *hierarchical decision structure*, was tested instead. The concept of a hierarchical decision structure is explained in section 11.1.2.

### 8.3 Lorenz

Lorenz specified two very similar models [Lorenz, 50 and 81] for action selection, both based on an analogy between *action specific potential* (the equivalent of drive strength for a particular system) and the pressure exerted on a valve by a body of water in a reservoir. Both models were designed mainly to account for ethological phenomena and address only a part of action selection, i.e. how to generate a drive strength for each system. The original model (figure 8.6) will be explained first, followed by a description of the changes made in the second model. Lorenz's ideas are more explicitly stated than Tinbergen's, but it is still only a "thought model", and is consequently rather vague in places.



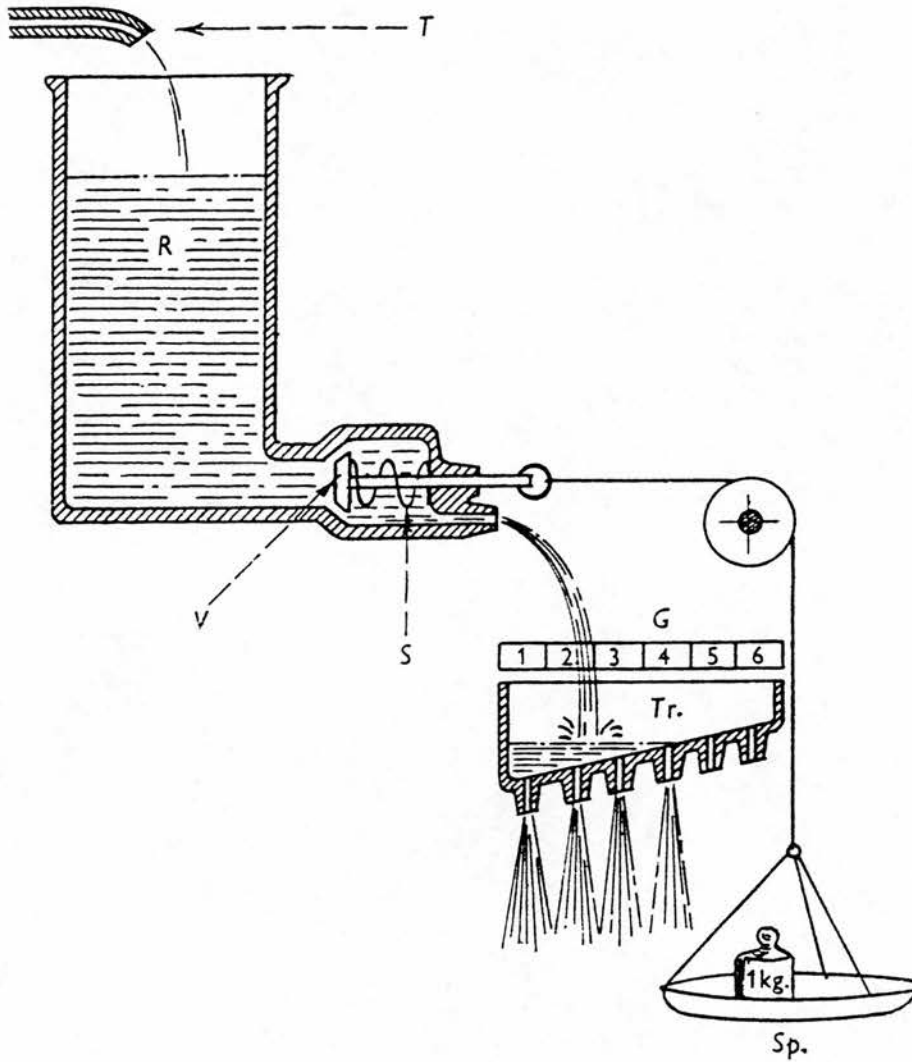


Figure 8.6: Lorenz's Original 'Psycho-Hydraulic' Thought Model. The labels are explained in the text. Taken from [Lorenz, 50].

In the diagram of the original "psycho-hydraulic model" (figure 8.6) the tendency for a type of behaviour to be expressed (its drive strength) is equal to the force being exerted on the valve  $V$ . This force is equal to  $(S_p + (k \times R) - S)$  where  $k \times R$  is the pressure which is proportional to the volume of water in the reservoir  $R$ ,  $S_p$  is the external stimulus strength (force on the scale pan), and  $S$  is the "inhibitory function of the higher centres". The water in the reservoir is the result of internal stimuli causing a continuous flow of water through the tap  $T$ . The rate at which water arrives in the trough  $Tr$  affects how full it gets and therefore also affects the number of outlets from



the trough through which water flows. The analogy is that the stronger the combined stimuli, the stronger the force on the valve and then the greater the volume of water that flows out per unit time. This then results in water coming out of more holes in the trough which corresponds to a more intense performance of the type of behaviour (e.g. to a more severe and daring attack on a rival). As the water comes out of the trough because the behaviour is being executed, the amount of water in the reservoir is reduced and so the pressure on the valve becomes lower.

One characteristic to be noted is that the internal variables are assumed to produce a constant input which is added cumulatively to the water reservoir. This implies that the type of behaviour will be released eventually even in the complete absence of appropriate external stimuli. There is evidence for such 'vacuum activity' in many animals. For instance, cats deprived of appropriate stimuli for a long time will stalk 'thin air'; fighting fish deprived of rivals to display at will eventually display at totally inappropriate objects or at nothing at all. This is the only model of action selection to be considered in which a constant value of an internal variable can produce increasing likelihood over time of the type of behaviour being expressed (i.e. of that system being chosen). Action specific potential (= water from the reservoir) is only ever removed when a consummatory action relevant to that system is executed.

Another important property of the model is that all stimuli (of whatever type) are added together. Again there is a "law of heterogeneous summation", this time for the stimuli activating the behaviour, rather than for the innate releasing mechanism as in Tinbergen's model.

The second, revised version of the model [Lorenz, 81], as shown in figure 8.7, is identical to the first except in the following ways: (i) additional readiness-releasing (i.e. 'priming') stimuli are added to the effect of the endogenous (internal) stimuli, and (ii) external stimuli are "added to the reservoir" instead of exerting a force on the valve via the scale pan. Lorenz also suggests that an element of inertia could be added to the opening and closing of the valve.

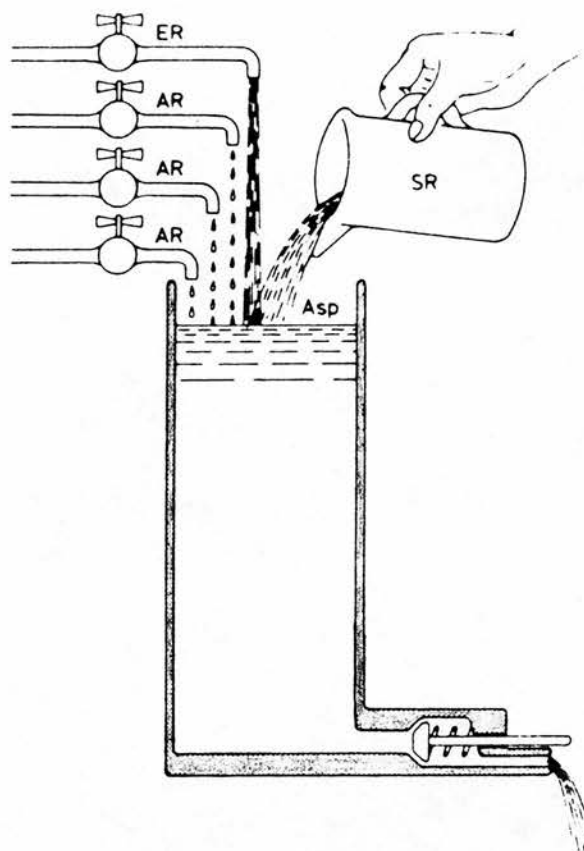


Figure 8.7: Lorenz's Revised 'Psycho-Hydraulic' Model. *ER* = endogenous or automatic stimuli, *AR* = 'unspecific readiness-releasing stimuli' and *SR* = 'specifically releasing key stimuli'. Taken from [Lorenz, 81].

---

Tinbergen and Lorenz's theories of action selection share an assumption of some energy or fluid which builds up and can only be 'got rid of' by execution of the right action. This assumption has been attacked [Hinde, 59] on the grounds that (i) the assumption is not required in order to explain the observed phenomena, and (ii) the assumption is not realistic. Only the computational properties of the model are scrutinised here.

Lorenz's model was implemented as an ASM and was tested in the SE. Specific details of the implementation are given in section 9.3.2.

## 8.4 Baerends

Baerends discussed the behaviour of two animals he had studied, the digger wasp and the herring gull, and then proposed mechanisms to account for aspects of their behaviour. These mechanisms are shown in figures 8.8 and 8.9. These two diagrams contain a lot of information, but a lot of the detail is not relevant to this discussion. The mechanisms are hierarchical (from left to right in the diagrams rather than top to bottom) with mutual inhibition between separate *systems*. Below the level of systems, each node chooses, in some undescribed fashion, which of its subordinates it will activate [Baerends, 76]. In the case of the herring gull, Baerends justified his set of hierarchies, one for each system or “higher-level instinct” (see figure 8.10), by claiming them to be the only way of explaining the observed patterns of behaviour. The entities furthest to the right are *actions*, or fixed action patterns (fixed action patterns are rigidly-ordered, commonly-occurring sequences of actions). The dotted lines in the diagrams represent feedback from the environment.

As with Tinbergen’s hierarchy, long-term decisions are incorporated as well as short-term ones. The mechanisms are only for two individual cases, rather than being a general prescription, but Baerends suggested the principles involved were generally applicable. Some aspects of the mechanics are not completely specified. It seems as if the effects of different internal and external stimuli are summed in all cases but this is not made explicit. The way in which higher-level nodes influence which of their subordinates are activated is similarly left unspecified.

Baerends calls his diagrams “functional explanations of behaviour”. They are not exact specifications of mechanisms and they do not contain many explicit or implicit assumptions about how actions are selected. Some points that Baerends does make are: (1) an ASM should be hierarchically organised, (2) higher-level nodes should affect or control which subordinates receive activation, (3) internal and external stimuli should affect which nodes at any level get chosen, and (4) once a node becomes active, it tends to stay active for some time.

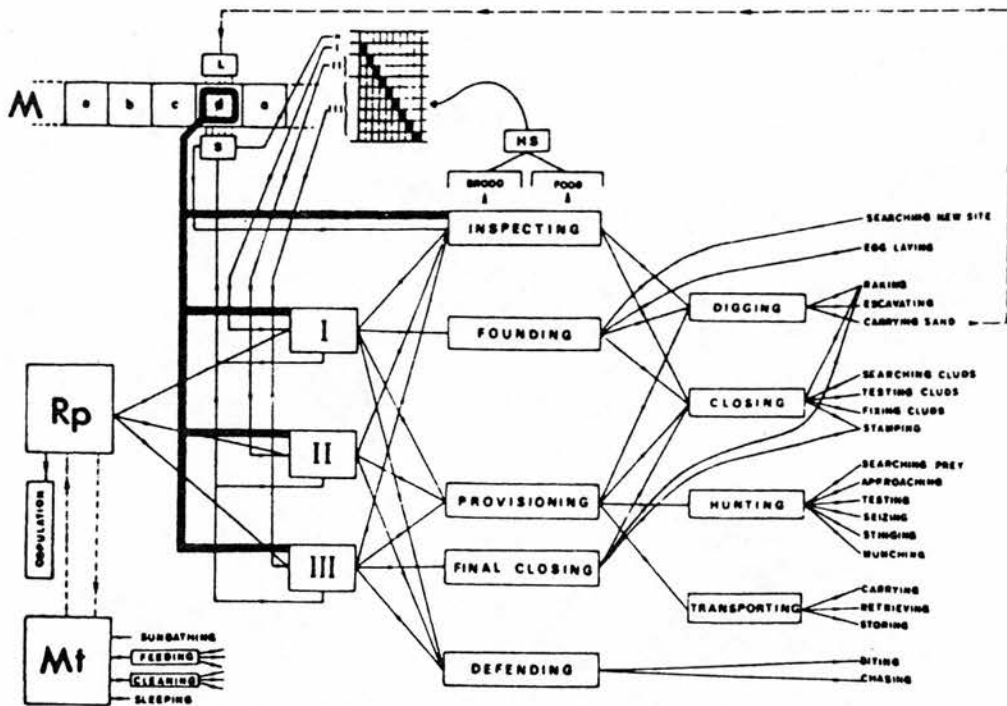


Figure 8.8: Baerends' model to account for nest provisioning behaviour of the digger wasp. As in the next diagram, actions or fixed action patterns are in the far right column and systems and sub-systems are to the left and middle. **Rp** = reproductive system, and **Mt** = 'maintenance' system. I, II and III are sub-systems controlling the completion of three separate phases of nest provisioning, each of which require the execution of several different lower-level sub-systems. Taken from [Baerends, 76].

Baerends' ideas were not implemented and tested, although the more generic hierarchical decision structure, mentioned under Tinbergen above, was (section 11.1.2).

## 8.5 Maes

One mechanism to come from the field of artificial intelligence is that by Maes [89, 90, 91a]. This is the most complex and the most completely specified of all of the mechanisms considered. It is a distributed, recurrent, non-hierarchical network. There are a

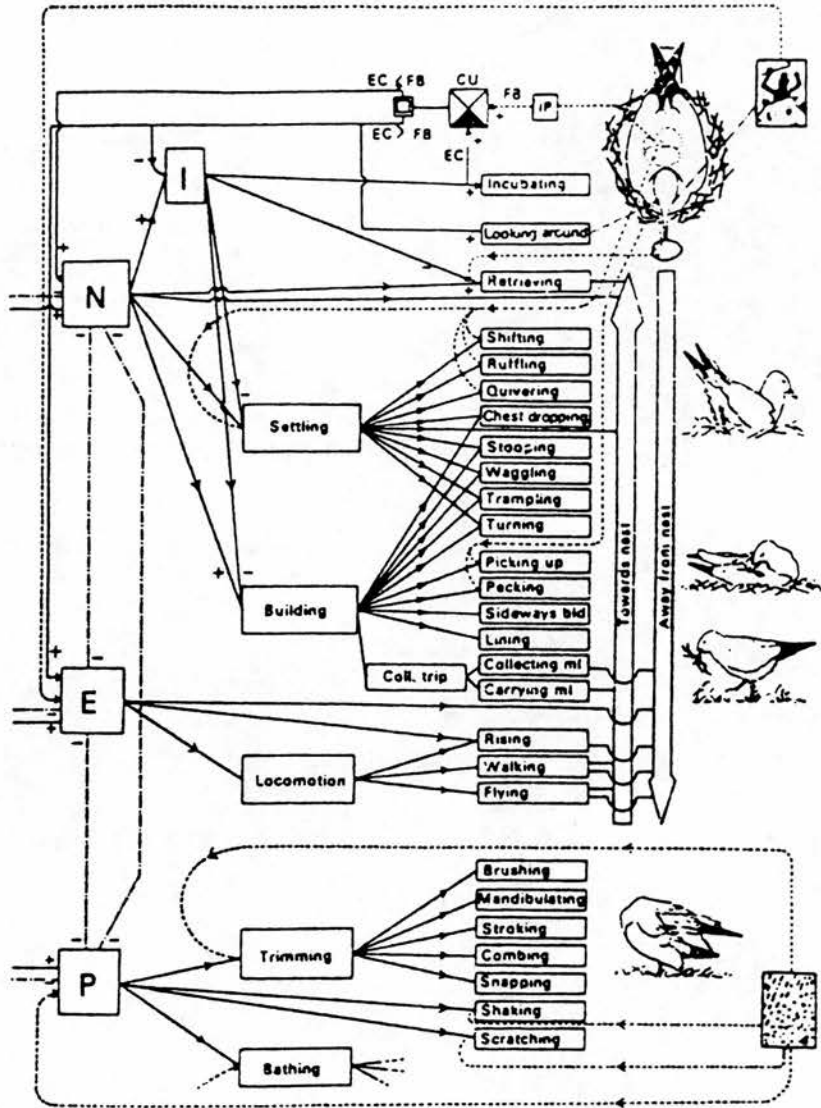


Figure 8.9: Baerends' model to account for incubation behaviour of the herring gull. Actions or fixed action patterns are in the far right column. "Superimposed control systems" of higher order are to the middle and left. N = incubation system, E = escape system and P = preening system. The three systems mutually suppress each other. Taken from [Baerends, 76].

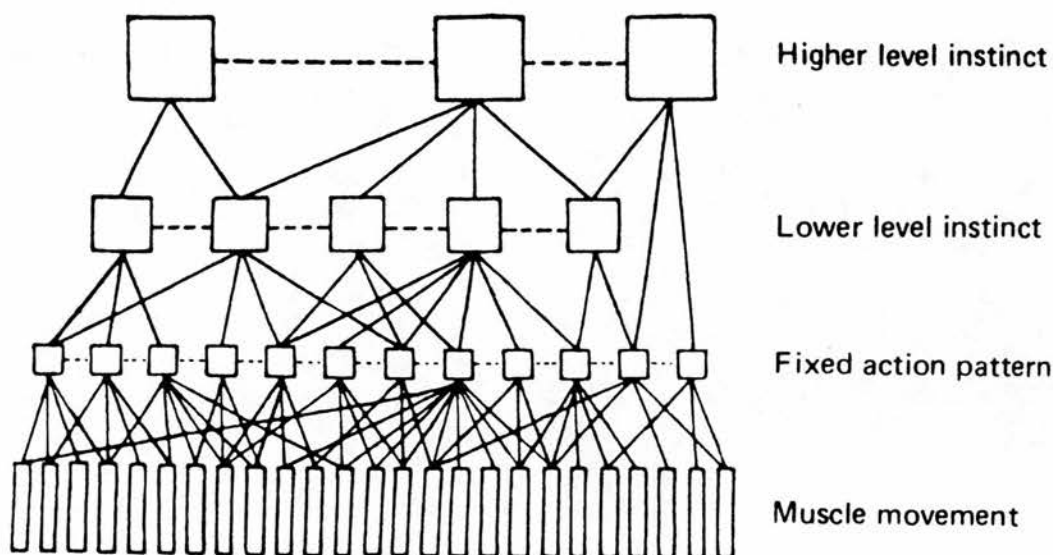


Figure 8.10: An example of the structure of a Baerends-like ASM. Dotted lines represent inhibitory relationships. Taken from [Lorenz, 81].

set of nodes which represent entities somewhere *below the system level* (e.g. approach food, drink water, flee from creature). The nodes are fairly simple, but are a lot more complex than neurons. The nodes can be either consummatory or appetitive.

There are two 'waves' of input to the network – firstly from the sensors of the environment (external stimuli) and secondly from motivations (usually derived from internal stimuli), as shown in figure 8.11. The sensors of the environment are binary-valued. That is to say, they measure the truth or falsity of logical statements about the environment. The motivations can be real-valued [Maes, 91a], and connect only to the consummatory nodes. There are also internal connections between the nodes which are of three types: predecessor, successor and conflictor links. After activation is passed in from the environment and the motivations then it is also spread around inside the network via the internal connections.

The characteristics of a node are now described more fully. Each node (see figure 8.12) has a set of preconditions. These preconditions are logical conditions about the environment which are required to be true in order for the node to be executable; i.e.



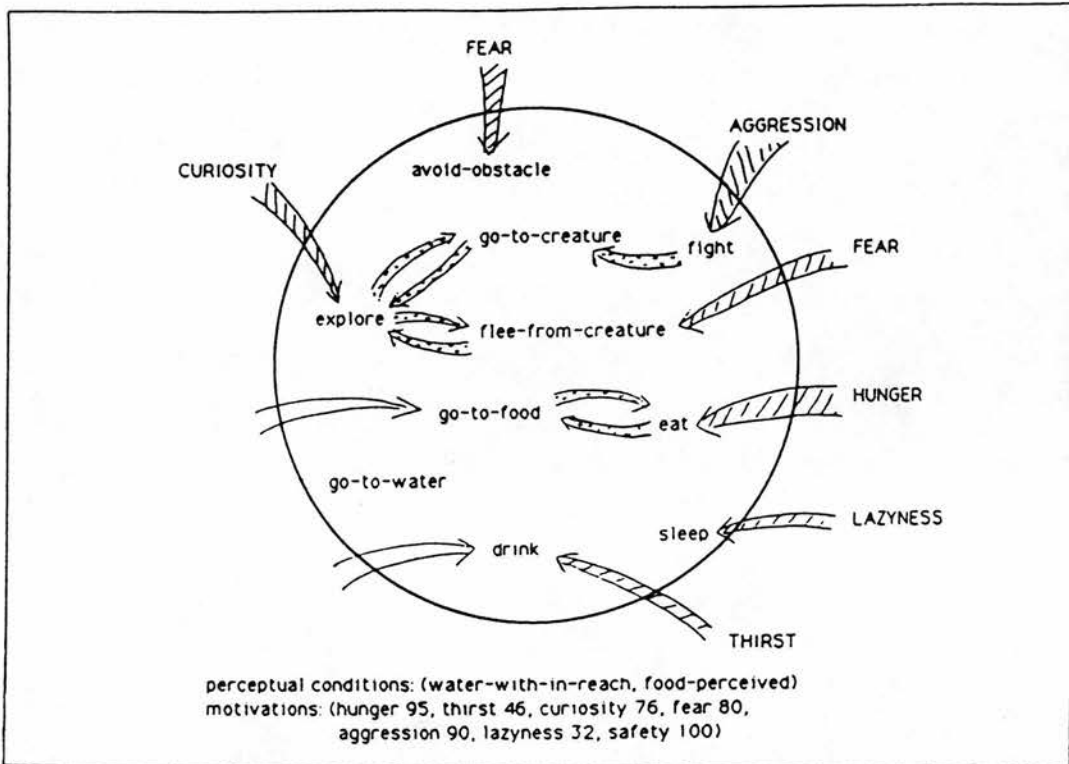


Figure 8.11: Maes' non-hierarchical "Bottom-Up Mechanism for Behaviour Selection". Taken from [Maes, 91a].

in order for it to be a valid, selectable alternative. For example, the precondition 'water in square' must be true in order for the node 'drink' to be executable. As well as a list of preconditions, each node also has an add list and a delete list. The add list consists of conditions about the environment that the node is likely to make true (e.g. 'water in square' for the node 'approach water'). The delete list consists of conditions that are likely to be made false by the execution of the node (e.g. 'no rivals nearby' for the node 'approach rival'). The final two components of the node are the activation level and the code which gets run if the node is executed (in some cases equivalent to a fixed action pattern).

The internal links in the network are specified as follows:

1. **predecessor links** – if (i) proposition  $X$  is false, (ii) proposition  $X$  is a precondition

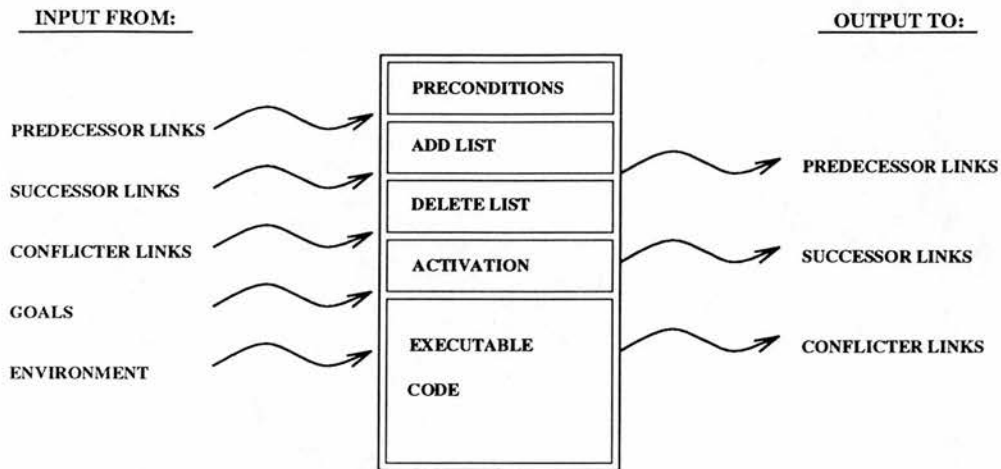


Figure 8.12: The make-up of a node in Maes' network.

of node *A*, and (iii) proposition *X* is in the add list of node *B* (i.e. if *B* can help *A* to become executable), then there is an active predecessor link (excitatory) from *A* to *B*.

2. **successor links** – if (i) proposition *X* is false, (ii) proposition *X* is in the add list of node *A*, (iii) proposition *X* is a precondition of node *B*, and (iv) the node *A* is executable (i.e. if *A* can help *B* to become executable), then there is an active successor link (excitatory) from *A* to *B*.
3. **conflictor links** – if (i) proposition *X* is true, (ii) proposition *X* is a precondition of node *A*, and (iii) proposition *X* is in the delete list of node *B* (i.e. if *B* stops *A* from becoming executable), then there is an active conflictor link (inhibitory) from *A* to *B*.

The external links providing input to the network are specified as follows:

1. **from sensors of the environment** – if (i) proposition *X* about the environment is true, and (ii) proposition *X* is in the precondition list of node *A* (i.e. if *A* is at



least partially appropriate to the current situation), then there is an active link (excitatory) from the sensor of the proposition  $X$  to node  $A$ .

2. **from goals** – if (i) goal  $Y$  has an activation greater than zero, and (ii) goal  $Y$  is in the add list of node  $A$  (i.e. if  $A$  is likely to achieve goal  $Y$ ), then there is an active link (excitatory) from the goal  $Y$  to node  $A$ .
3. **from protected goals** – if (i) goal  $Y$  has an activation greater than zero, and (ii) goal  $Y$  is in the delete list of node  $A$  (i.e. if  $A$  is likely to undo goal  $Y$ , or stop it from being achieved), then there is an active link (inhibitory) from the goal  $Y$  to node  $A$ .

The exact procedure used to select a node to execute at each timestep is as follows (see figure 8.13):

1. calculate the excitation coming in from the environment and the motivations.
2. spread excitation along the predecessor, successor and conflictor links.
3. check to see if any nodes are executable, and if so then choose the one with the highest activation, execute it and finish.
4. otherwise, if no node is executable, reduce the global threshold and repeat the cycle.

A node is executable if all of its preconditions are true and if its activation is greater than the global threshold. If more than one node is executable after a cycle then the one with the highest activation is chosen. When a node has been chosen and executed then its activation is reset to zero before the next timestep. All other node activations are normalised so that the average of activations becomes equal to the constant  $\pi$  (see below).

Several global parameters can be used to 'tune' the performance of the ASM to a particular environment:

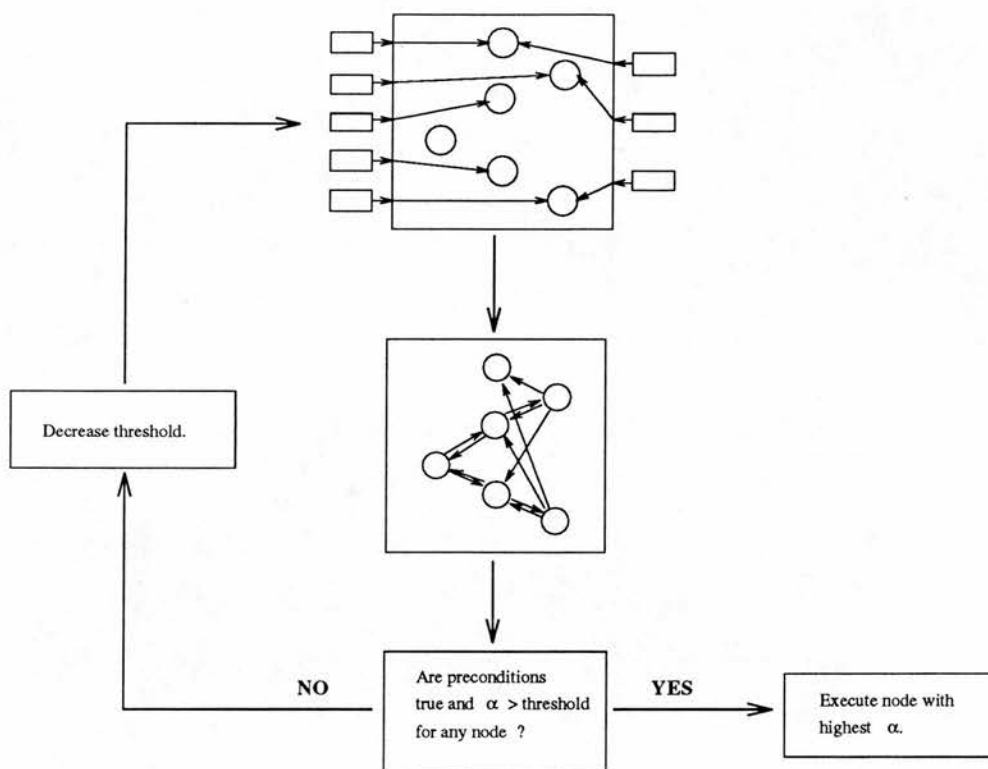


Figure 8.13: The procedure carried out at each timestep in Maes' mechanism.

1. the mean activation value after each timestep (used in normalisation) [ $\pi$ ].
2. the initial value of the global threshold [ $\theta$ ], which is reduced by 10% after each cycle if no executable node has an activation greater than it.
3. a constant determining the weighting of environmental sensor inputs (as opposed to other sorts of inputs), as well as the weighting of successor links (as opposed to other sorts of links) [ $\phi$ ].
4. a constant determining the weighting of goal inputs and predecessor links [ $\gamma$ ].
5. a constant determining the weighting of protected goal inputs and conflictor links [ $\delta$ ].

The different inputs to a node are multiplied by the following: (i) environmental sensors –  $\phi$ , (ii) goals –  $\gamma$ , (iii) protected goals –  $\delta$ , (iv) successor links –  $\frac{\phi}{\gamma}$ , (v) predecessor links – nothing, and (vi) conflictor links –  $\frac{\delta}{\gamma}$ .

There are also division rules affecting how much activation gets added or taken away from nodes as a result of the different sorts of links in the network. Consider the situation in figure 8.14a, in which there is a predecessor link from node  $B$  to node  $A$ , due to a proposition  $X$  being in both the precondition list of  $B$  and the add list of  $A$ . Node  $A$  has  $M$  elements in its add list (one of which is proposition  $X$ ) and there are  $N$  different nodes that can make  $X$  true (i.e. that receive predecessor links from  $B$  because of proposition  $X$ ). In this case the predecessor link increments the activation of  $A$  by an amount  $\frac{\alpha}{NM}$ , where  $\alpha$  is the activation of node  $B$ .

Similar division rules are prescribed for the successor and conflictor links, and also for the inputs to the network from outside it. For instance, when activation is fed in from a sensor of the environment due to a proposition being true (see figure 8.14b), then node  $A$ 's activation is again incremented by  $\phi \frac{\alpha}{NM}$ , where in this case  $\alpha = 1$  (propositions are either true or false),  $N$  is the number of nodes that have proposition  $X$  in their precondition list, and  $M$  is the number of preconditions of node  $A$ .

The effects of internal and external stimuli (mediated through internal connections for some nodes) are always summed. The mechanism has nodes at only one level, which is below the system level but in some cases above that of the action level. The only distinction between nodes is that between appetitive and consummatory. Consummatory nodes receive direct motivational input, whereas the appetitive nodes only receive that input indirectly through the consummatory nodes or through other appetitive nodes closer in the sequence to the consummatory node. All nodes can receive direct input from sensors of the state of the environment.

The central idea of Maes' model for action selection is that the different types of links encode various relationships (e.g. consummatory / appetitive relationships be-

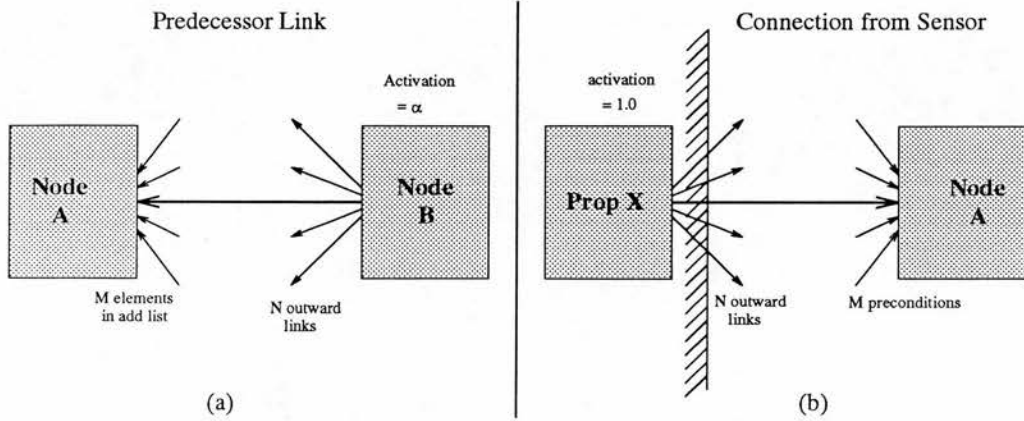


Figure 8.14: The division rules applied to excitation transferred along (a) predecessor links, and (b) inputs from the environment. In (a) the activation of A is incremented by  $\frac{\alpha}{NM}$ ; in (b) the activation of A is incremented by  $\phi \frac{\alpha}{NM}$ .

tween nodes, conflict relationships between nodes, goal-achieving relationships between nodes and goals, goal-counteracting relationships between nodes and goals and situation-dependency relationships between environmental sensors and nodes), and with this information hard-wired into the mechanism, excitation can just be fed in from the current situation and current goals and activity will, after several iterations, come to settle in the node representing the most appropriate behaviour.

Maes' mechanism is more complex than the others. This is partly because it is specified more precisely than the other mechanisms and so there is more detail to consider. Maes' ASM was one of those that were implemented and tested in the SE. It was chosen both because of the lack of ambiguity in the description and also because it is radically different from the other ASMs considered.

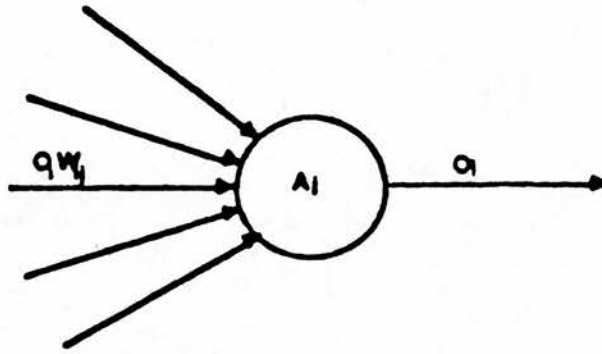
## 8.6 Rosenblatt & Payton

This mechanism was conceived by two robotocists [Rosenblatt & Payton, 89]. It is a connectionist, hierarchical, feed-forward network. It was designed partly in response to the work of Brooks (see section 8.8.1). Rosenblatt & Payton do not give any examples of mechanisms to solve whole action selection problems. They only give examples of parts of a whole ASM that are suitable for solving parts of the action selection problem for an autonomous robot. As such it is not always easy to know how exactly they would decide to apply their ideas to the particular problem posed by the SE, which is designed to reproduce animal-like, rather than robot-like decision problems. In some cases their ideas have been extrapolated in order to produce a complete ASM to deal with the demands made by the SE. These cases are pointed out in the text.

Their suggested type of mechanism is similar in many ways to the hierarchies of Tinbergen and Baerends except that more than one node is allowed to be active at any one level. They do not specify any intra-layer inhibition or competition, and indeed this would be counter to their assumption of combination of preferences, to be explained below. The only competition is at the level of the behavioural final common path, necessarily so, where a winner-take-all process must occur. The ideas are more akin to Baerends' than to Tinbergen's. There are none of Tinbergen's IRMs (stimuli only affect node activations) and no stipulation that appetitive actions are generated by higher-level nodes than those which generate consummatory actions.

Figure 8.15 shows the makeup of a node in this sort of network. The node is a fairly standard artificial neuron except in that the node activation can be *any* function of the weighted inputs, not necessarily just a weighted sum.

Figure 8.16 gives an example of a Rosenblatt & Payton network. There is a hierarchy of nodes which receive information from internal, indeterminate and external stimuli and feed their activations down through the hierarchy until they arrive in the *action* nodes, i.e. those at the level of the behavioural final common path. At that level a



$O_i$  is the output of unit  $i$ .

$W_{ij}$  is the weight on the link from unit  $i$  to unit  $j$ .

$A_j = f_A(O_1 W_{1j}, \dots, O_n W_{nj})$ , is the activation level of unit  $j$  with  $n$  weighted inputs.

$O_j = f_O(A_j)$ , is the output of unit  $j$ .

Figure 8.15: A node from Rosenblatt & Payton's hierarchy. Taken from [Rosenblatt & Payton, 89].

winner-take-all process decides which action should get executed. The way in which different stimuli should be combined is left unspecified (because the node activation can be any function of the weighted inputs).

The hierarchies in figure 8.10 (Baerends) and figure 8.16 (Rosenblatt & Payton) appear to be almost identical. Inputs from stimuli have been omitted from the first diagram, but there is no difference in their effect: they increase a node's activation in both cases and there are no IRMs. The main difference between the two sets of ideas, which is not apparent from the two diagrams, is that the Rosenblatt & Payton scheme allows for *combination of preferences* (see figure 8.17). Instead of one node at every level being selected (through a process of intra-layer inhibition or whatever), and only that one being allowed to remain active, Rosenblatt & Payton allow all nodes to remain active in any layer. Instead of only one node in each layer being able to excite other nodes below it in the hierarchy, Rosenblatt & Payton allow each node to remain active and *express a degree of preference* for each of the nodes beneath it (i.e. further down in the hierarchy). The preferences of nodes for subordinate nodes can be positive or negative. All the preferences for a node are combined according to some rule, which is not necessarily

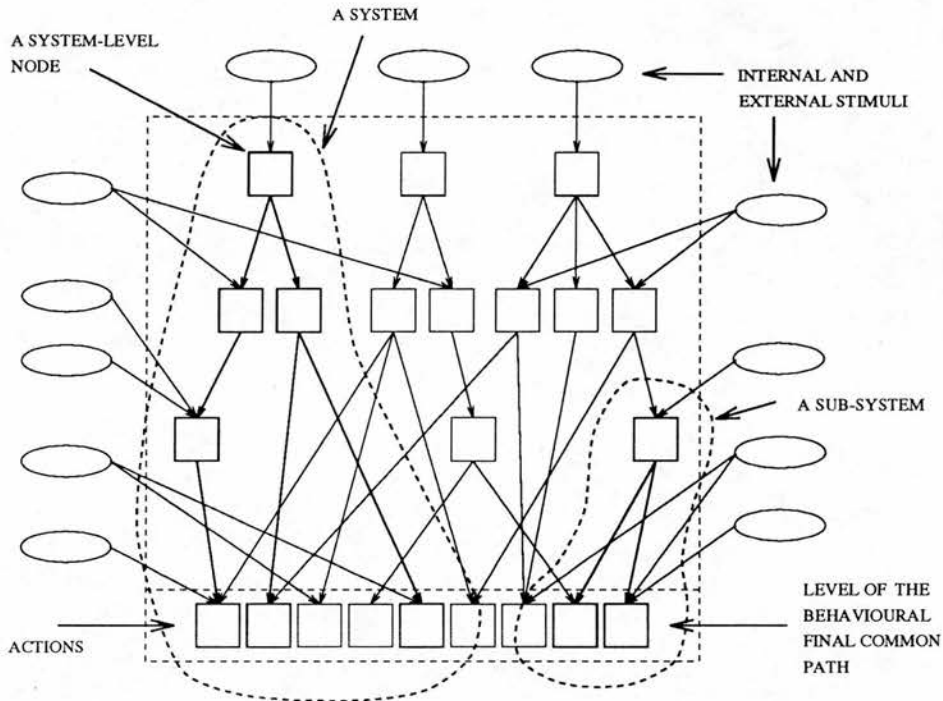


Figure 8.16: An example of the structure of a Rosenblatt & Payton-like ASM. A single system is shown outlined by the curved dashed line. Stimuli are also shown impinging on the mechanism.

a simple summation.

The Rosenblatt & Payton ASM is not explicit about all aspects of action selection (how *should* inputs be combined? should there be differential treatment of appetitive and consummatory alternatives?), but those aspects that are defined are fairly unambiguous. It represents a different type of hierarchy from the more conventional ones of Tinbergen & Baerends and for this reason it was chosen as one of the ASMs to implement and test in the SE.

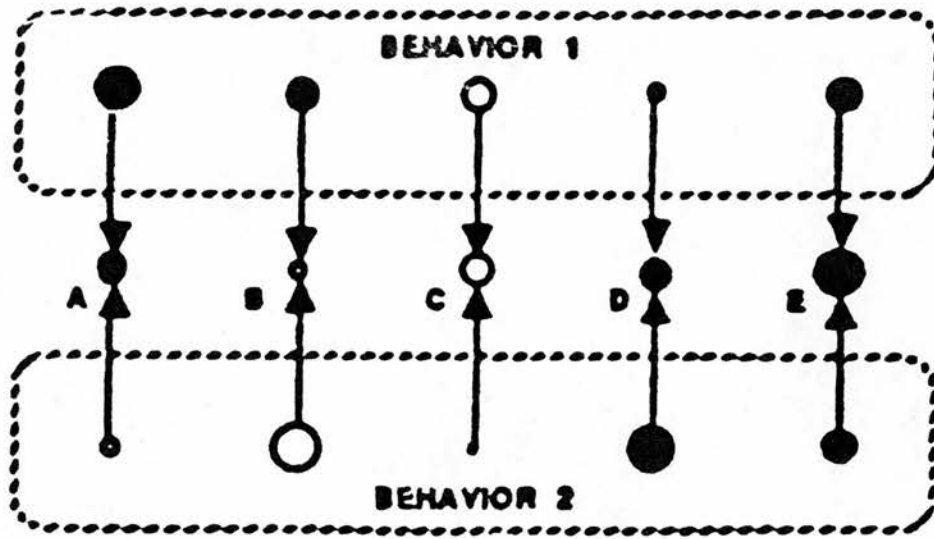


Figure 8.17: Combining evidence in a Rosenblatt & Payton hierarchy. Taken from [Rosenblatt & Payton, 89].

## 8.7 Halperin

Halperin's mechanism [Halperin, 91, and 92] was designed, first and foremost, to account for aggressive display behaviour in Siamese fighting fish. The mechanism has one central component, a novel synaptic modification rule (the "neuro-connector rule") that can account for many different conditioning phenomena. Since the synaptic modification rule enables learning it is ignored here. Halperin also describes the structure of the mechanism within which such learning must take place. It is this aspect of the work which is of interest here.

The mechanism (see figure 8.18) is a hierarchical, feed-forward, connectionist network. Only two different levels in the hierarchy ('motivational' or system, and 'behavioural' or action) are mentioned, but other levels are not ruled out. The entities at the different levels are similar, but only those at the behavioural (action) level connect to motor control areas. There is a winner-take-all competitive inhibition between entities at the action level. There is also some weaker inhibition between entities at the system level. This does not effect a winner-take-all operation, but rather is expected to produce some sort of selective attention. The hierarchy is therefore not a "rigid switching mechanism"



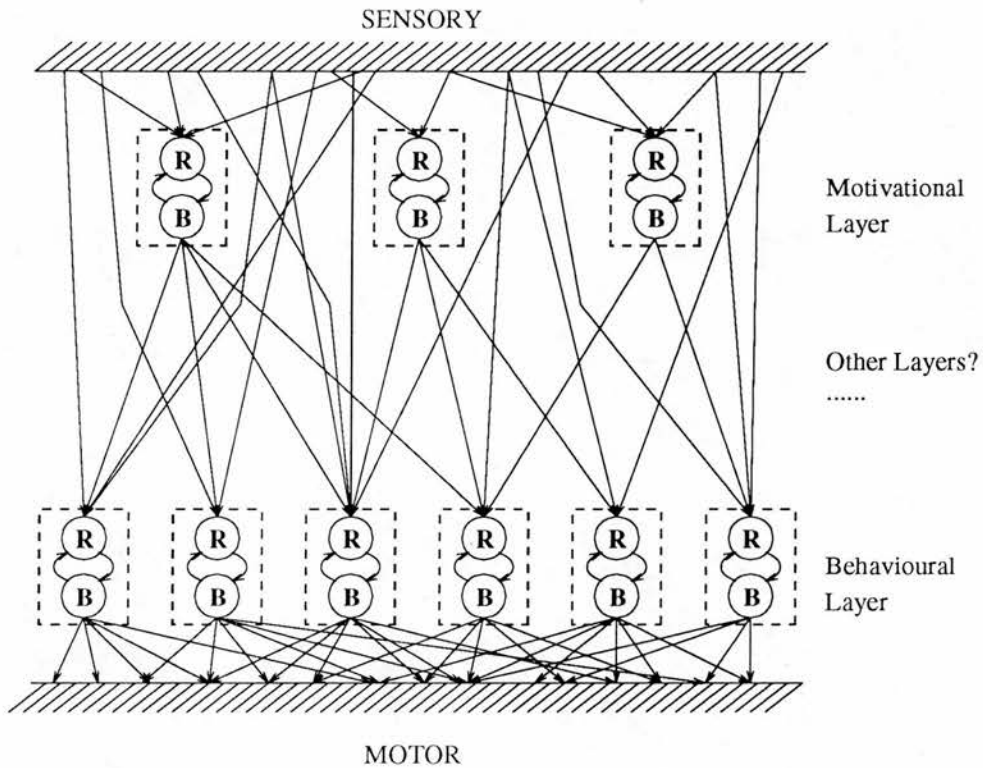


Figure 8.18: The structure of Halperin's ASM. *R* = releaser nodes, *B* = behaviour nodes. There is positive feedback between the *R* and *B* nodes.

in the same way that Tinbergen's or Baerends' hierarchies are. It is possible for mixed motivations to be expressed.

The *R* (releaser) nodes in figure 8.18 act to *sum* the relevant stimuli for the *B* (behaviour) nodes. There is also a positive feedback loop between *R* and *B*. The *B* neurons in the motivational layer provide inputs to the *R* neurons in the behavioural layer. These inputs are treated as normal stimuli and summed with other stimuli in calculating the activation of the *R* neurons in the behavioural layer. It is stated that the entities in the behavioural layer will generally require both ordinary stimuli and motivational stimuli from above (the motivational layer) in order for them to become active.

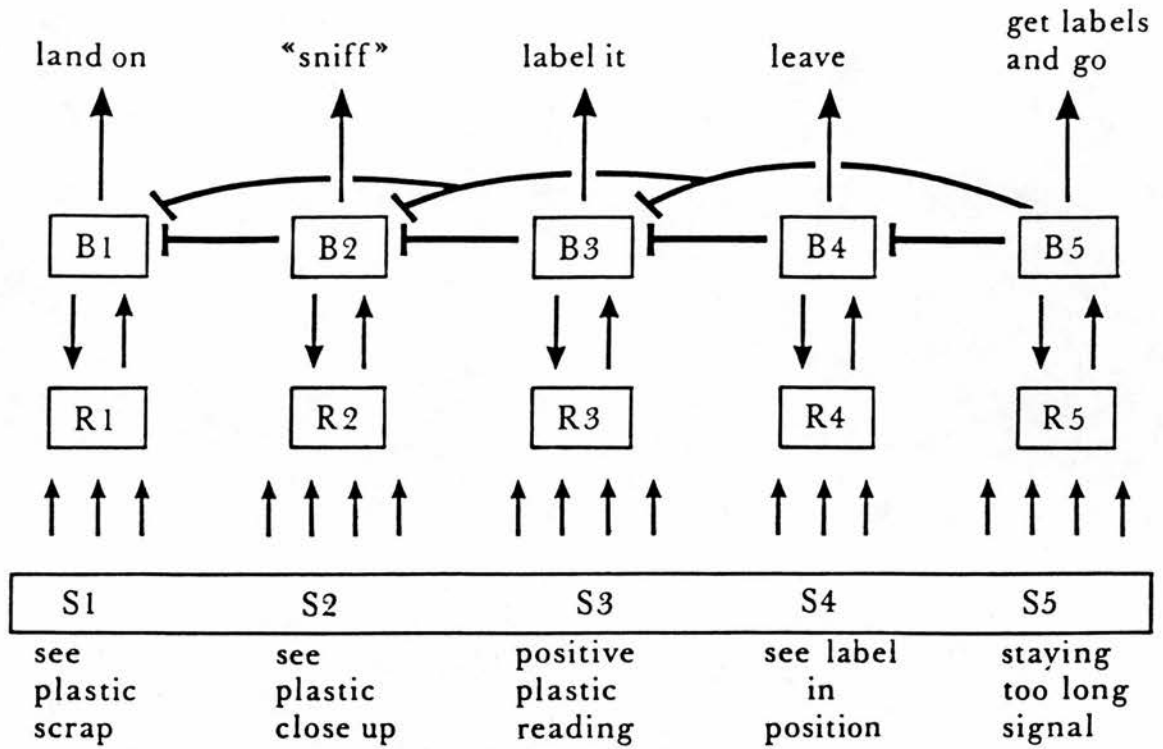


Figure 8.19: An imaginary ‘scrap-labelling’ robot using an instance of Halperin’s ASM. Taken from [Halperin, 91].

Figure 8.19 shows a set of behavioural entities from an example instance of the mechanism designed to control the behaviour of a hypothetical “scrap-labelling” robot. The robot is normally intended to perform the sequence *B1* through to *B4* unless *S5* becomes active, in which case *B5* interrupts and takes over from whichever of *B1* – *B4* was being executed.

This mechanism was not implemented and tested, partly due to the lack of precision of some aspects of the description, at least for some of the non-learning aspects of the mechanism. For instance, how should the partial inhibition between motivational entities be implemented? And what strength of positive feedback should there be from a *B* to an *R* node?

## 8.8 Other Related Work

All the mechanisms to be considered in this thesis have now been described. This section now looks at other work which is more loosely related to mechanisms for action selection and discusses any insights to be gained from the work.

### 8.8.1. Brooks

One noticeable omission from the previous list of ASMs is the *subsumption architecture*. This was proposed by Brooks [86, and 87b] as an approach to computing action selection and motor control, and perhaps perception to some degree as well, in robots. As mentioned briefly in chapter 1, Brooks believes that intelligence should not involve central planners or overriding control processes. Instead, intelligence should be comprised of many distributed processes which are only able to act coherently together. Each separate process, or *competence module*, a set of finite state machines in his case, should receive its own, unprocessed copy of whatever parts of the sensory input it is interested in, rather than receiving an interpretation of the environment made by some other module. Each competence module works independently of every other one and continuously generates output signals as a function of input signals unless *inhibited* or *subsumed* by another module.

An example of the subsumption architecture is shown in figure 8.20. The different types of interaction that are allowed between modules are shown in figure 8.21.

Each competence module (task-achieving entity) is a collection of finite state machines. All the lowest-level competence modules achieve fairly simple tasks such as avoiding obstacles, and are designed independently of each other or of higher-level modules. The higher-level modules can send signals to effectors, but can also inhibit or subsume the activity of other lower-level modules. To subsume the output of a target module

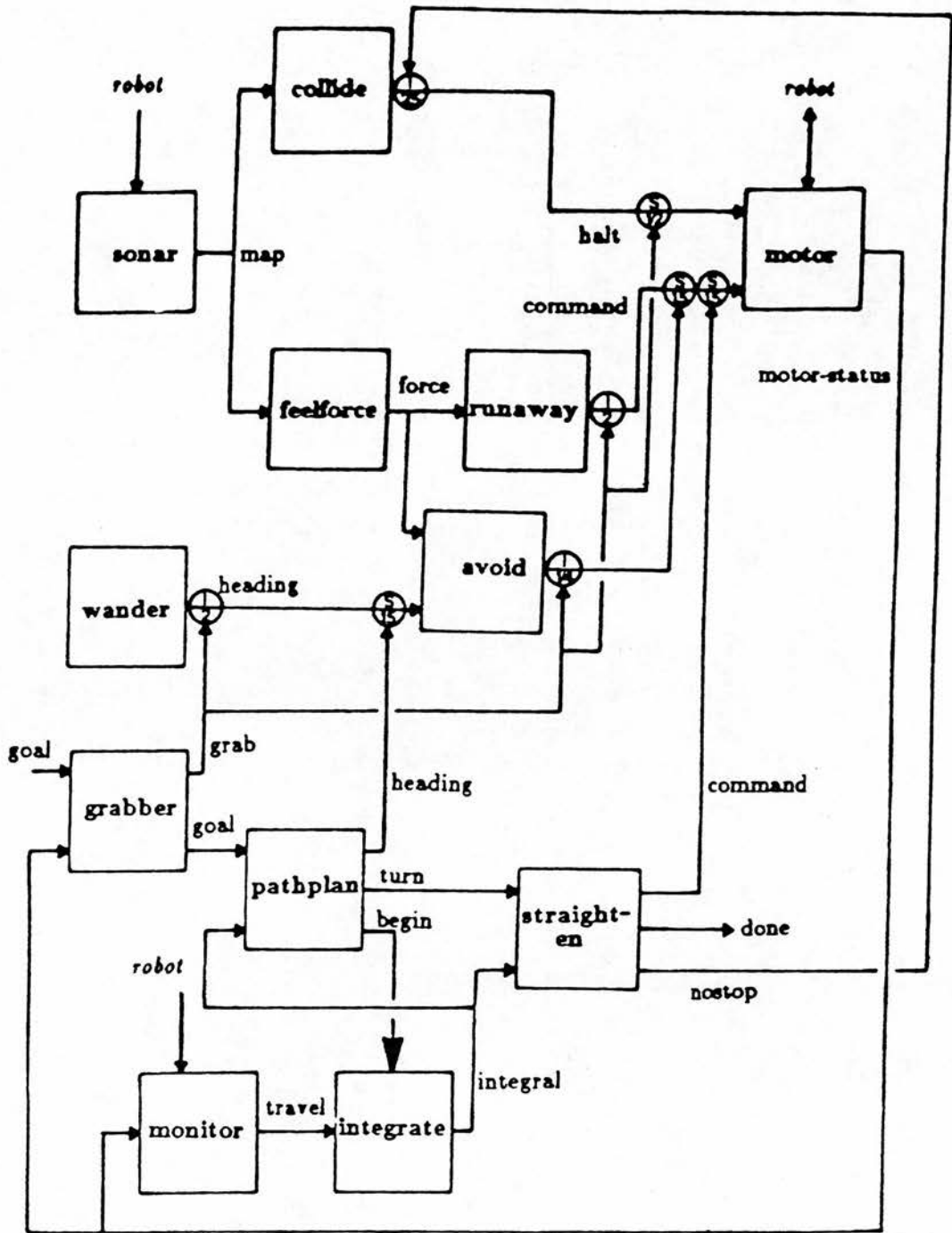


Figure 8.20: An example of the subsumption architecture. A circle with an S at the top indicates subsumption, a circle with an I at the top indicates inhibition. Taken from [Brooks, 86].

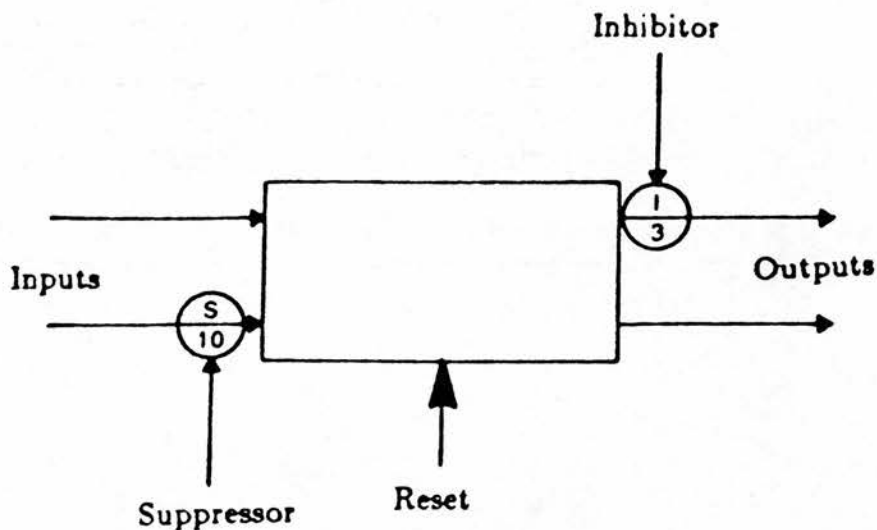


Figure 8.21: A module can affect another in several different ways: (i) it can provide an input signal to it, (ii) it can reset its state, (iii) it can subsume one of its outputs or (iv) it can inhibit one of its outputs. Taken from [Brooks, 86].

---

means to replace its output with another value. That is, to inhibit the output of the target module *and* at the same time to overlay another signal in its place.

The reason that this formulation is not included in the list of ASMs is that it is not strictly an action selection mechanism, but rather an *architecture* or *computational substrate* which can be used as a means of instantiating many different ASMs, or indeed as a means of instantiating different techniques for motor control. To put it in terms of Marr's levels of analysis [Marr, 82], the subsumption architecture is neither a computational description of the problem of action selection, nor an algorithmic description of how to select actions, but rather it is an implementational technique, which can be used to implement any number of different algorithms. The subsumption architecture does not contain a description of how to deal with important aspects of the action selection problem such as how stimuli should be combined, or how different systems or actions should be arbitrated between (or even if they should).

The subsumption architecture has proved to be a successful technique for implementing the control of robots, as it was intended to be. However, some of the advantages of

the technique, for instance the ease with which new competence modules can be added to the existing ones without too much reconfiguration, are more relevant to the task of building robots than the task of selecting actions. In short, it is an implementational scheme rather than an algorithm or mechanism for selecting actions.

### 8.8.2. Ludlow

Ludlow [80] presented a simple neural network that is capable of performing a winner-take-all operation between alternative systems. It only tackles a small part of the AS problem, that of how to choose the system or action with the largest activation in a neurobiologically plausible manner. It doesn't consider different levels, or how to combine stimuli, or how to calculate the 'importance variable' (drive strength) for each system or action – only how to perform the winner-take-all, or 'maximum selecting' operation once all the variables have been calculated.

The mechanism (see figure 8.22) works in the following way: if one of two neurons initially receives more input than the other then it fires and inhibits the other which eventually ceases to fire (assuming the weights on the inhibitory links are greater than or equal to 1.0). If the inhibited neuron later starts to receive more input from stimuli than the currently firing one then it will overcome its inhibition and start to inhibit the currently dominant neuron. After several cycles the second neuron will have taken over and become dominant over the first, which will be fully inhibited and will have ceased to fire. If the inhibitory links have weights which are greater than 1.0 then the system will always tend to a state in which only one of the two neurons is active, if it is not in that state already. Weights which are significantly greater than 1.0 will cause persistence since an active neuron will be able to remain active even when its inputs are slightly less than the dominated neuron.

These arguments can be extended to a similar system with more than two neurons.

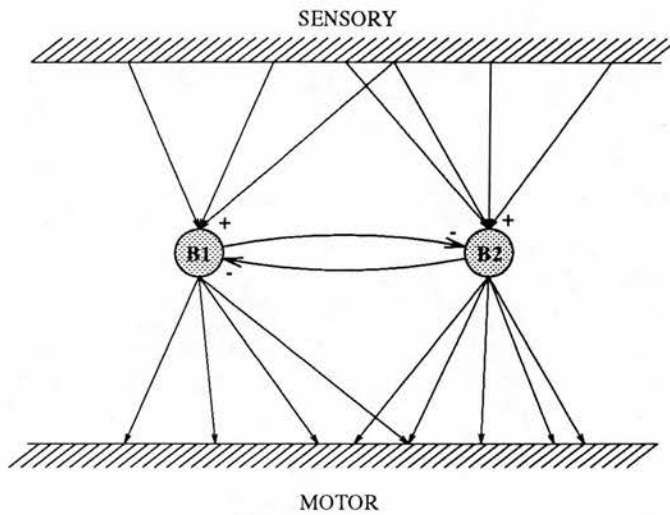


Figure 8.22: Ludlow's "Decision Maker". *B1* and *B2* mutually inhibit each other (after [Ludlow, 80]).

---

### 8.8.3. Beer

Beer has looked in detail at the locomotory and feeding behaviour of an artificial insect [Beer, 90; Beer, Chiel & Sterling, 90; Beer & Chiel, 91]. The insect was modelled in the computer and tests were carried out in a simple simulated environment. The work is mainly concerned with motor control, but does touch briefly upon the subject of action selection. Figure 8.23 shows the action selection mechanism for the artificial cockroach. There are several notable features: (i) there is no hierarchy, although there are dominance relations mediated by inhibitory connections, and (ii) the motivational stimulus of 'feeding arousal' (i.e. internal food deficit) excites both consummatory and appetitive alternatives directly. This is in contrast to Maes' ASM, in which motivational stimuli only excite consummatory nodes directly, and Tinbergen's hierarchy, in which motivational stimuli impinge directly only upon the higher-level nodes, which are responsible for the less immediate appetitive actions.

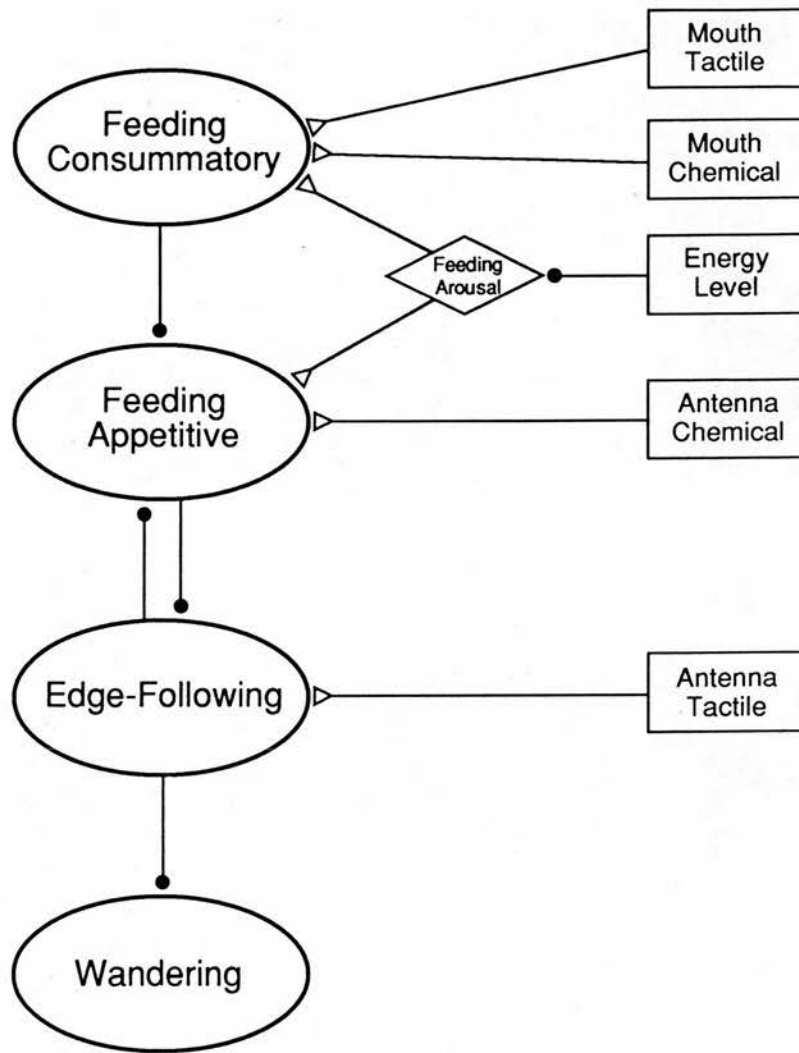


Figure 8.23: Action selection mechanism for Beer's cockroach. Taken from [Beer, 90].

#### 8.8.4. Planning

Planning is a more traditional way of thinking about the problem of action selection, especially among researchers in the field of artificial intelligence. The essence of the difference between typical planning systems and the ASMs already described is that planning systems are *deliberative*, whereas the mechanisms here are *reactive*. In general, planning systems work by comparing the current situation to some goal, and then constructing a *plan*, a sequence of actions, that will take the animal from its current



state to the goal state. The idea is that the agent then follows that plan blindly to its end, after which it will have arrived at its goal state, and can look around again to decide upon another goal and another plan. In contrast, the systems just described in sections 8.1–8.7, while they do possess goals (e.g. ‘reduce food deficit’, ‘escape from a predator’), sense their internal and external environment continuously, and recalculate their chosen action continuously instead of following a plan which may have been formulated sometime in the past.

Planning systems are more deliberative partly because the sorts of problems that they have been tested on are more predictable and deterministic. For instance, some of the problems that have been, or are being, tackled by planning systems are: (i) how to make cups of coffee, (ii) how to navigate from *A* to *B* around known obstacles, (iii) how to assemble components in the right order, and (iv) how to rearrange stacks of blocks (by moving one block at a time from one stack to another).

Classical planning, with its pre-planned sequences of actions and its assumptions about the predictability and stability of the outside world, has been heavily criticised by many authors [e.g. Agre & Chapman, 87, Agre 90, Maes, 90a, Beer, 90]. The main criticisms are as follows:

1. Since (i) the animal will not have perfect knowledge of the world, and (ii) changes to the world will occur which have not been brought about by the animal, then any plan is likely to need reformulating frequently because the world will not behave as predicted (for instance, an unexpected hazard will occur between the animal and the food source it is heading towards, or the fruit food source that has been approached will not be in the right part of its ripening-decaying cycle and so there will be no food there). A realistic environment is not completely deterministic or stable from the point of view of the animal, and so fixed plans do not usually work well.

2. Plans need to be interruptible, to take account of urgent situations or beneficial opportunities (e.g. while heading towards a water source, the animal sees a predator, or encounters a high value food source).
3. Many aspects of planning are by nature combinatorially explosive and computationally intractable in complex environments. For instance, the number of possible states is enormous (of the order of  $10^{100}$  for the SE).
4. Planning systems are centralised and non-distributed. They assume a central location in the system which interprets all of the perception of the animal and uses the result to produce a plan. This leads to a central bottleneck and a consequent lack of robustness.

Some more recent approaches to planning [e.g. Firby, 87] have taken into account some of the criticisms of planning and are more reactive. Firby takes the "extreme position of using no prediction of future states at all", and his system recalculates its plan after every timestep, instead of following an old one blindly.

#### **8.8.5. Learning Systems**

Another related field of research is the study of how best to learn what actions to select in different situations from trial and error experience in an initially unknown environment. Systems such as those proposed by Sutton & Barto [e.g. Sutton, 88], Watkins [1989], and Booker [1988] are *tabula rasa* ('blank slate') mechanisms with no predispositions towards certain actions in certain situations. By trying random actions in different situations and taking note of the degrees of success of those actions, the animal can gradually build up a mapping between situations and appropriate actions. The issues of interest in this field are, for example, how to assign credit to appropriate actions, how to assign credit to all actions in a necessary sequence (not just the consummatory action) and how to set the balance between exploitation (using the knowledge already obtained) and exploration (to obtain more knowledge).

There are three main reasons why learning systems have not been implemented and evaluated in the SE:

1. *tabula rasa* systems with no hard-wired innate tendencies would fare very badly in this SE. In many cases it is unlikely that the animal would be able to survive the necessary tests to discover that, for instance, it does not pay to try and mate with predators or to approach dangerous place features. Even in less critical situations, an animal which instinctively knows to eat cereal type food will have a distinct advantage over one which has to discover it.
2. In an SE of this complexity, the possible number of situations is enormous (the size of the state space is of the order of  $10^{100}$ , even with fairly coarse quantisations of continuous variables). The number of possible mappings between situations and actions that has to be explored is enormous, and is not feasible in an individual animal's lifetime.
3. In order to get interesting results in the time available it was necessary to restrict the subject of investigation to the hard-wired inflexible aspects of action selection, rather than trying to study learning as well.

One valuable concept in this field [Sutton, 88], is that of *temporal discounting*. In this scheme, the likely reward  $R$  of any action, or sequence of actions, is multiplied by a *temporal discount factor*  $\alpha^t$ , where  $\alpha$  is less than 1.0 and  $t$  is the number of timesteps that will probably be required to obtain the reward. So, if a strategy will obtain a reward (increase in expected genetic fitness) of 0.5, but it will take 10 timesteps to achieve it, then the discounted reward is equal to  $(0.5 \times \alpha^{10})$ . By comparing discounted rewards, action sequences which achieve smaller rewards, but in a shorter time-scale, can be preferred over those which achieve larger rewards over a much longer time.

### 8.8.6. Neurophysiology of ASMs for Simple Animals

The physiology of some simple animals is fairly well understood and the way in which action selection is implemented can be explained. For instance some bacteria [Koshland, 1979] have only a very few actions such as ingest food, split in two to reproduce, move forwards and tumble to orientate in a new random direction. They have sensors for changes in the intensity of beneficial and detrimental chemicals and changes in temperature. They control their movement towards or away from these by altering their likelihood of tumbling according to whether the concentration of the relevant chemicals (or the temperature gradient) is increasing or decreasing as they move forwards.

Molluscs such as *Pleurobranchaea* and *Aplysia* have very simple nervous systems and behaviour, both of which have been extensively studied [e.g. Davis *et al* 74, Davis, 79, Kovac & Davis, 80, Kandel, 79]. There appears to be fairly strict dominance relationships between different behaviours, not dependent on the values of motivational stimuli such as food deficit, as shown for *Pleurobranchaea* in figure 8.24.

Although the explanation of how real physiological systems accomplish action selection in these simple animals is interesting, the understanding of these processes in molluscs is not at all complete, and for higher animals is a lot more rudimentary. The actual physiology of real animals does not help greatly in constraining or increasing our understanding of action selection mechanisms. At present, the lessons that have been learnt from 'bottom-up' approaches of this sort do not give much insight into how to construct action selection mechanisms for more complex problem environments.

## 8.9 Summary of Differences Between Mechanisms

Some of the important differences between the seven ASMs described in sections 8.1–8.7 are shown in table 8.1.

	Drives	Tinbergen	Lorenz	Baerends	Maes	R & P	Halperin
Network?	×	√	×	√	√	√	√
Hierarchy?	×	√	×	√	×	√	√
System-Level Competition?	√	√	√	√	×	×	partial
Combination of Preferences?	×	×	×	×	×	√	unstated
Method of Arbitration Between Options in Same System	unstated	IRMs and competition on basis of activation	unstated	selection by higher level nodes in hierarchy	via predecessor & successor links	×	partial competition
Combination of Stimuli	unstated	arbitrary	summed	unstated	summed	arbitrary	summed
Effect of Unvarying Internal Stimulus	constant effect	constant effect	increasing effect	constant effect	constant effect	constant effect	constant effect
Sensors of External Environment	unstated	real-valued	real-valued	real-valued	binary-valued	real-valued	real-valued

Table 8.1: Structural differences between ASMs and differences in how they deal with certain aspects of the action selection problem.

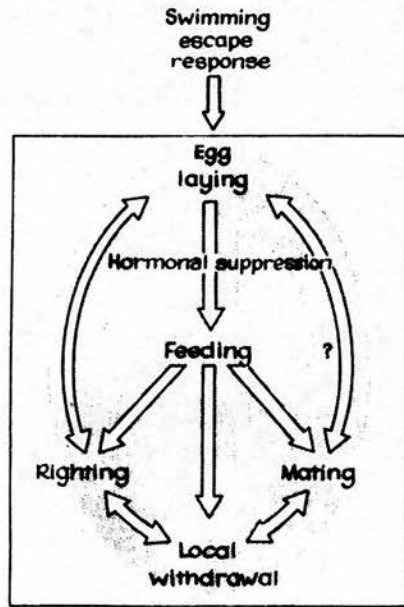


Figure 8.24: Dominance relationships between behaviours for *Pleurobranchaea*. Taken from [Davis *et al*, 74].

---

## Chapter 9

# Performance of Mechanisms

Due to time constraints only a subset of the ASMs considered in the previous section were implemented and tested in the SE. The drives model was implemented first because it was the simplest. Lorenz, Maes and Rosenblatt & Payton, and a hierarchical decision structure (see section 11.1.2) were also chosen to give a good variation in types of mechanism (e.g. symbolic *v.* connectionist and hierarchical *v.* non-hierarchical).

This chapter describes how these different mechanisms were tested and presents the results obtained. Where the SE exposed obvious deficiencies in the performances of mechanisms then the underlying causes of the deficiencies were investigated, and remedied where this did not involve too substantial a change to the original proposed form of the mechanism. However, in cases where the deficit in performance was due to a problem with a central and intrinsic aspect of the ASM then no modification was made.

### 9.1 How Each Mechanism was Tested

#### 9.1.1. Test Procedure

The procedure for evaluating each mechanism in the SE was as follows:

1. **implementation:** the descriptions of the mechanisms were used to specify the implementation of them. Any ambiguities or gaps in the specifications were noted and resolved sensibly.
2. **preliminary testing:** the performance of each ASM was observed directly in the SE, using the graphical displays showing (i) the animal in the SE, (ii) what the animal perceives of the SE, (iii) the animal's internal variables and (iv) the animal's 'map' of the environment. Any obvious deficiencies (e.g. never moving, never performing consummatory actions, never performing complete sequences, not avoiding hazards, not avoiding predators) were noted.
3. **adjustments to improve performance:** any gross deficiencies in the performances of the mechanisms were analysed using the tools described in section 3.6 and their root causes identified. In most cases the problems could be overcome by small changes to the mechanism, by changes in the representation of the input or by changes in the parameter values of the mechanism. Occasionally though, fixing a shortcoming required a change to the central structure or dynamics of the mechanism. If the change required was fairly minor then it was carried out. If, however, a fairly radical change to the fundamental structure or workings of the mechanism was required (e.g. to introduce hierarchical aspects to Maes' ASM) then this was not carried out.
4. **fine-tuning by parameter adjustment:** if it had proved possible to remove all gross shortcomings in the performance of a mechanism by simple changes, then the next step involved fine-tuning the mechanism by altering the parameter values so as to try and maximise the average genetic fitness produced. Approximately equal time was spent on this fine-tuning for all ASMs.
5. **obtaining a result:** once the parameter values of the ASM had been finalised then its performance was measured in the SE by carrying out about 1650 tests, each with a different, random initial set-up of the SE (see chapter 4). The performance is the non-inclusive genetic fitness measure, i.e. the number of times that the animal manages to mate before it dies (see section 2.3). The *average* over the 1650



or so tests was taken as the measure of the performance of the ASM.

6. **checking for bias in the SE:** tests were also carried out in three other SEs which had been altered in different ways. The different SEs were:

- (i) the nature of the model of perception was changed by altering the effect of the time of day on the effectiveness of perception, by increasing the obscuring effect of intervening vegetation, and by changing the effect of different actions on the efficacy of perception.
- (ii) the nature of the model of navigation was changed by increasing  $\sigma^2$  (the variance of the estimate of how far the animal moves each timestep) by a factor of 3.0, and by reducing the size of the 'map'.
- (iii) the nature of the model of motor control was changed by increasing the likelihood of executing an action incorrectly, and by changing the effects of different actions on the conspicuousness of the animal, and on the animal's body temperature and food and water levels.

The purpose of carrying out the tests again in these different SEs was to check for any bias in the results because of the particular models of the processes of perception, navigation and motor control in the SE (i.e. to check that an ASM gets a better result because it is able to select actions more appropriately, rather than because it is better suited to the individual characterisations of the processes of perception, navigation and motor control incorporated in the SE).

### 9.1.2. Ensuring Equal Treatment of Each ASM

There are two types of result generated by these tests in the SE. Firstly, there is the detection of gross inadequacies in the performance of an ASM, which are then traced back to infelicities in the ASMs design. Secondly, there are performance measurements for different ASMs which can be compared with each other. In general, instances of the first sort of result are fairly hard proofs that the design of the ASM is sub-optimal for

action selection. Results of the second sort (comparisons between ASM performances) are less secure, as discussed in section 5.3. Part of the reason for this is that it proved impossible to use genetic algorithms to optimise the parameters in each mechanism.

In order to ensure the fairest possible comparisons between ASMs, even though these comparisons will never constitute more than fairly weak evidence anyway, the following steps were taken during testing: (i) as explained above, each ASM was modified as necessary so that small design faults were not allowed to give the impression that the whole mechanism was imperfect. (ii) on the other hand, no ASM was changed beyond recognition. Therefore what was finally measured in the SE always kept the essential features of the original specification. (iii) approximately equal time was given to 'parameter-tweaking' for each ASM, so that differences in performance were unlikely to be solely due to differential amounts of fine-tuning of the ASM. (iv) equation 4.2 (section 4.3) was used when necessary to calculate the statistical degree of certainty of a real difference in performance in the SE for two ASMs.

## 9.2 Factors Affecting the Desirability of Different Alternatives

The function of an action selection mechanism is to select, at each moment in time, an action for the animal which maximises (or at least comes close to maximising) its *future expected genetic fitness* (section 2.3). It was argued in section 5.5 that, for several non-trivial reasons, it is impossible to calculate the exact future expected genetic fitness for each action in each state of a realistically complex environment, but yet despite this each mechanism has to select between actions somehow.

In all of the mechanisms there are variables that correspond to the desirability of actions, sub-systems or systems. For instance, the drives ASM has one variable (drive strength) per system and in Baerends' ASM there is one variable (node activation) for each action, sub-system or system. There is an implicit assumption for each ASM that the values of the variables for each option are in some way approximately proportional to the fitness of carrying out that option. But in practice it is impossible to calculate exact

values of the future expected genetic fitness and so the different ASMs necessarily use different methods for constructing *approximations* to the fitness of different alternatives. The drive strengths, node activations or other variables that represent the tendencies for an alternative to be chosen should rise or fall in accordance with several more immediate and available factors or indicators of the future expected genetic fitness. These factors, which are enumerated and discussed below, should be represented in some way, either implicitly or explicitly, in the decision-making mechanism. This is not to say that somewhere in every animal's brain there will always be explicit neural signals corresponding to such factors, but rather that these factors need to be taken into account in some way in an action selection mechanism that is computationally optimal.

### **9.2.1. Expected Short-Term Change in Health**

What is the expected increase or decrease in the animal's health if this action is or is not carried out now? For instance, if the animal's health is currently reduced because of lack of water then drinking water will have an immediately beneficial effect. Being attacked by a predator, or encountering a hazard may also result in an immediate, detrimental change to the animal's health.

### **9.2.2. Expected Long-Term Change in Health**

What is the expected long-term increase or decrease in the animal's health if this action is carried out now? For instance, exploring will probably increase health in the long term because the animal will find out more about food and water sources and will be less likely to suffer from health decrements from lack of food and water. Also, moving back towards the den is likely to reduce the variance of the animal's estimate of where it is, make it less likely to get lost, make it more likely to be able to find its den again before nightfall, and thus reduce the chance of it dying due to encounters with predators or hazards at night.

### 9.2.3. Expected Short-Term Risk of Death

As the animal's health gets closer to zero then it becomes more and more likely that the animal will die soon (see figure 9.1). Small perturbations in health (e.g. from a temporary drop in body temperature because of a drop in external temperature) become increasingly likely to kill the animal. For this reason the animal will, if its health is very low, prefer to choose a smaller, but more immediate increase in health over a greater, but more delayed increase. On the other hand, if its health is less critical then it may choose the latter option.

In addition, imagine the case where the animal's health is reduced due to both lack of food and lack of water, with lack of food causing the larger decrement. Imagine also that the animal can increase its health by a larger amount per timestep by eating food than by drinking water. In this situation it may still be beneficial to the animal to choose drinking before eating, if the time left before the animal dies due to lack of water alone is less than the time left until the animal dies due to lack of food alone (see figure 9.2). To put it another way, if water will 'run out' quicker than food then the animal may prefer to try and get water rather than food, even though this will result in a smaller immediate increment in health.

This factor has a lot in common with the concept of 'risk-space' and the idea that the behaviour of the animal may be arranged so as to maximise its distance from lethal boundaries [McFarland & Houston, 81].

### 9.2.4. Direct Effects on Genetic Fitness

The previous three factors are all important because they impinge upon the animal's ability to survive, but survival is only important in so far as it allows the animal to reproduce in the future. Actions such as courting and mating have no effect on the animal's health but are desirable because they can lead to direct increases in the animal's genetic fitness.

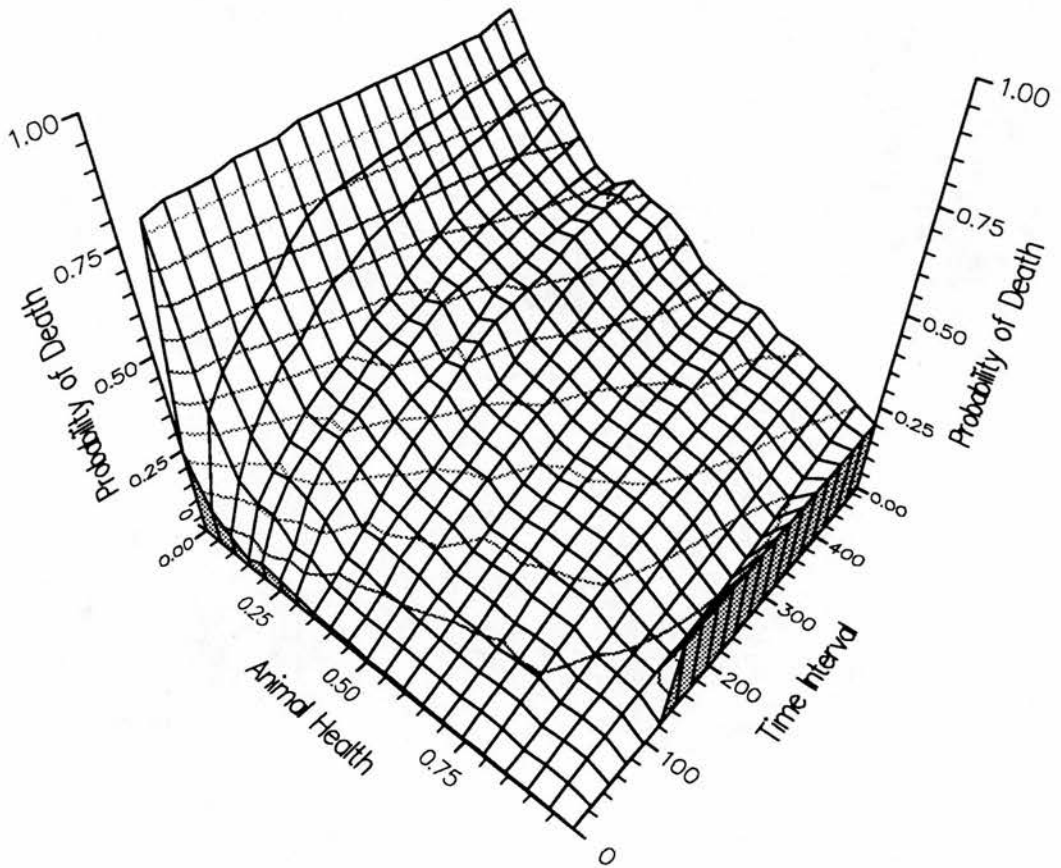


Figure 9.1: Graph showing the probability of death over different intervals of time from the current moment in time, for different values of the animal's health at the current moment in time. The animal was using the drives ASM to select its actions. It can be seen that the probability of death does not decrease linearly with increasing animal health. Note that averaging errors mean that (i) the probability of death is not always shown as 0.0 for a time interval of 0.0, and (ii) the probability of death is not always shown as 1.0 when the animal's health is equal to 0.0. In addition, the small number of results for values of animal health greater than 0.9 (it is unusual for the animal to be able to achieve such a high degree of health) make the graph less smooth in that region.

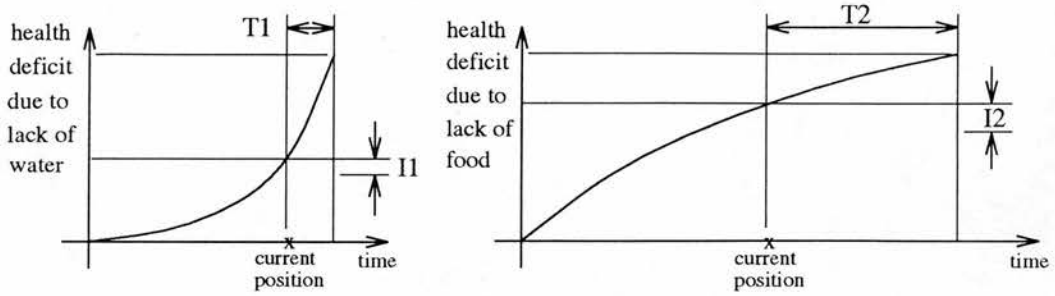


Figure 9.2: Hypothetical situation in which obtaining water should be chosen in preference to obtaining food, even though the food deficit is larger and the animal can reduce more of it in a single timestep.  $T1$  and  $T2$  are the times it will take for the animal to die from that deficit alone if no rectifying behaviour is undertaken.  $I1$  and  $I2$  are the respective increases in health that can be gained from the execution of one timestep of consummatory behaviour.

### 9.2.5. Taking Advantage of Opportunities

How *time-critical* are the benefits of a particular action? For instance, if the animal is likely to derive a smaller increase in health from pouncing on (i.e. eating) some prey, than from cleaning itself, then it may still be preferable to catch the prey simply because the opportunity to do so does not arise very frequently, whereas the animal can groom or preen itself whenever it wants to. In general, if when the animal performs action  $A$  followed by action  $B$  it gets benefits from both, but when it performs action  $B$  followed by action  $A$  it only gets benefit from action  $B$ , then it may pay the animal to perform  $A$  followed by  $B$ , even if the benefit from  $B$  is greater than that from  $A$ .

### **9.2.6. Continuing Expenditure of Time Required to Obtain Benefit**

If an option does not directly give rise to an immediate or delayed benefit itself, but only when one or more other options follow it, then how much extra time must be expended on the remaining options in order to realise the benefit? For instance, an appetitive action must be followed by a consummatory one in order for it to have been worthwhile. In general, if option A performed by itself yields a benefit  $R$  (no matter whether that benefit is delayed or immediate) then it is more valuable than an action B, which is part of a sequence BCD that yields the same benefit  $R$  after all of B, C and D have been performed.

## **9.3 Implementational Issues and Performance Results**

This section now presents the results of testing several ASMs in the SE described in Part I. For each ASM, the details of its implementation are listed, followed by descriptions of any serious inadequacies in its ability to select appropriate actions, and resolutions to these inadequacies where possible, as well as performance results for the final version of each ASM.

### **9.3.1. Performance of Drives ASM**

#### **First Implementation**

The first drives mechanism to be implemented was the most naïve one; i.e. the one in which the drive for 'get water', for instance, was derived using only the internal stimulus 'water deficit'. In general, no external stimuli were used to calculate the drives strengths for systems unless there were no relevant internal or indeterminate stimuli. Some example drive strengths from this initial implementation were calculated as follows:



1. **get water** –

$$drive\_strength = 0.95 \times water\_deficit$$

where *water\_deficit* is the distance below the satiation level for water. If the animal's health is low (less than 0.25) then an extra 0.1 is added to the drive strength. The total drive strength is not allowed to exceed 0.95.

2. **escape predator1s** – if the predator1 is perceived with a high degree of certainty and is less than 3 squares away then

$$drive\_strength = 0.99,$$

otherwise if the predator1 is a bit further away or the perception of the predator1 is a bit less certain to be correct then

$$drive\_strength = 0.40 \times predator1\_stimulus.$$

Finally, if no predator1 is perceived, or one is perceived fairly uncertainly and at a distance, then

$$drive\_strength = 0.00$$

3. **avoid hazards** – if the dangerous place is perceived with a high degree of certainty and is in an adjacent square then

$$drive\_strength = 0.98$$

Otherwise if the dangerous place is a bit further away or the perception of the dangerous place is a bit less certain to be correct then

$$drive\_strength = dangerous\_place\_stimulus$$

Finally, if no dangerous place is perceived, or one is perceived fairly uncertainly and at a distance, then

$$drive\_strength = 0.00$$

4. **reproduce** –

$$drive\_strength = 0.80$$



5. **sleep in den at night** –

$$drive\_strength = 0.95 \times night\_proximity$$

where *night\_proximity* is a measure of how close it is to nightfall. This indeterminate stimulus has a value of 1.0 during the night, which drops to 0.0 at sunrise, remains close to zero through the day, and then increases towards 1.0 as sunset approaches.

6. **scan for predators** –

$$drive\_strength = 0.90 \times time\_since\_scanned$$

where *time\_since\_scanned* varies in the range 0.0 (animal chose a look action in the last timestep) to 1.0 (animal hasn't chosen a look action in the last 6 timesteps). There can also be a further increment to the drive strength of up to 0.5 if the animal has perceived a predator recently, and more recently than the last look action. The total drive strength is constrained to a maximum of 0.96.

As explained in section 9.1, the equations and numerical values used here were set by hand and then improved by trial and error. Similar equations were used to calculate all of the other drive strengths in this initial implementation. As described in section 8.1, the system with the highest drive strength wins the competition and an action is chosen that best suits the needs of that system alone. Some examples of the rules that were used to decide the most appropriate action for a system were as follows:

1. **get water** – if there is a reasonable stimulus value for water in the same square then choose the action DRINK. Otherwise, if there is a fairly high stimulus for perceived water, and that stimulus is greater than the one for remembered water, then choose a MOVE action so as to approach the perceived water. Otherwise, if there is a fairly high stimulus for remembered water, and that stimulus is greater than the one for perceived water, then choose a MOVE action so as to approach the remembered water. Finally, if the animal cannot perceive or remember any decent water sources then explore (choose a semi-random direction to move in).

2. **escape predators** – if the perception of the predator is fairly uncertain (there is a high probability of it being ‘imaginary’) and the perceived predator is not very close then look in that direction. Otherwise, if the perception is more certain or the predator is perceived to be fairly close, then if the animal is in its den just rest (it is invulnerable to attack there), otherwise if it is in fairly dense protective vegetation then ‘freeze’ (remain completely motionless) to try and avoid detection. Otherwise, if the animal is not in good cover but can perceive cover then it should move towards it. If no proper protection can be perceived then the animal should just run in the opposite direction from the predator.
3. **avoid hazards** – move in the opposite direction from the dangerous place, unless it is in the same square in which case move in a random direction.
4. **reproduce** – if a courted mate is perceived in the same square then mate with it. Otherwise if a receptive mate is perceived in the same square then perform a courtship display towards it. If no receptive or courted mate is perceived in the same square but a receptive mate is perceived fairly strongly in another square then approach it. Finally, if there are no reasonably strong perceptions of mates in other squares then explore.
5. **sleep in den at night** – if in the den then sleep, otherwise if the den is perceived then approach it. If the den cannot be perceived then approach the remembered position of the den.
6. **scan for predators** – if there is a significant perception of either type of predator in any direction then look in that direction, otherwise just look around generally (choose the action LOOK\_AROUND).

The first, basic implementation of the drives ASM has now been described. When tested in the SE it performed fairly well, although one expected fault was apparent. As discussed in section 8.1, systems such as ‘get food’ and ‘get water’ should include some contribution from external stimuli in the calculation of their drive strengths. In this naïve calculation in the first implementation there was no opportunism – the animal

just travelled over a good food or water source unless that happened to be the winning system anyway because of the respective internal deficit. Some sort of dithering was also observed, and there was a noticeable tendency towards lack of completion of action sequences because the drive strength did not increase as the animal got closer to the consummatory action in an appetitive-consummatory sequence (e.g. as the animal approached a mate, courted it and then copulated with it).

## Second Implementation

To remedy these faults, the drive strengths for the systems 'get food', 'get water', 'warm up', 'cool down' and 'reproduce' were changed to incorporate external stimuli. These changes gave rise to the second implementation. Two examples systems from it are:

### 1. get water –

$$drive\_strength = (0.95 \times (water\_deficit + (water\_deficit \times water\_stimulus)))$$

where *water\_stimulus* is the maximum of the stimulus for perceived water and remembered water. There is still a 0.1 increment for low health and a maximum of 0.95. Note that the drive strength cannot be proportional to  $(water\_deficit + water\_stimulus)$  alone since this would be greater than zero even if there was no deficit (if the animal was satiated). Also, the drive strength could not be proportional to  $(water\_deficit \times water\_stimulus)$  alone since this will be zero when *water\_stimulus* is zero, even if the animal is almost dead due to lack of water.

### 2. reproduce –

$$drive\_strength = 0.30 + (0.62 \times mate\_stimulus)$$

where *mate\_stimulus* is 1.0 with a courted mate in the same square, close to 1.0 with a receptive mate in the same square and somewhat less for receptive mates in other squares.

This second version of the drives mechanism was implemented and observed in the SE. The problems with dithering and lack of contiguous action sequences no longer occurred. There were no obvious major deficits in the performance of the ASM. McFarland [1985, pp 461] hypothesised that simple motivational competition (drives) might not be satisfactory because low-priority systems such as grooming could never win the competition. This did not prove to be a problem here because the functions that calculated the drive strengths were adjusted so that each system won the competition with approximately the desired frequency.

When this second and final version was evaluated in the four different versions of the SE then the average results ( $\bar{x}$ ) were 6.44, 3.29, 6.41 and 8.78, with an average over the four of **6.23**. The standard deviations ( $s$ ) of the individual measurements, in the tests in each of the four versions of the SE, were 6.42, 3.89, 6.55 and 8.17; and for the average result over all four versions was 6.26. The expected standard deviations of the sample averages ( $s_{\bar{x}} = (s/\sqrt{n})$ ) were 0.16, 0.09, 0.16 and 0.20; and for the average result over all four versions was 0.15.

### **9.3.2. Performance of Lorenz's ASM**

As described in section 8.3, Lorenz's model is somewhat similar to the drives model in that an drive strength is calculated for each system, although in this case the *action specific potential* (as the equivalent of drive strength is called by Lorenz) is equivalent to the resultant force on the valve in figure 8.6, obtained from summing the force from the scale pan (external stimuli) and the force from the head of water in the reservoir (internal stimuli). An identical procedure to the drives model was used for selecting an action for the winning system, it was only the method of calculating the drive strength, or action specific potential that was different.

## First Implementation

The initial, naïve implementation was as follows: all internal stimuli exerted their influence by controlling the flow of water through a tap into the reservoir, all external stimuli exerted their influence as weights which could be put on and taken off from the scale pan, and all indeterminate stimuli were assigned to either taps into the reservoir or weights on the scale pan as appropriate. Therefore the influence from an internal stimulus (water from the tap) remained behind even if the internal stimulus value returned to zero, whereas the influence of an external stimulus (weight on the scale pan) only existed as long as the external stimulus did. The reservoirs were emptied only when a consummatory action for the system was performed. Some example systems are:

### 1. get food –

$$\text{new\_reservoir\_force} = \text{old\_reservoir\_force} + (2.2 \times \text{food\_deficit})$$

at every timestep, where *food\_deficit* is the maximum of the deficits for fat, carbohydrate and protein. And

$$\begin{aligned} \text{scale\_pan\_force} = & (10.0 \times \text{perceived\_food\_in\_same\_square\_stimulus}) + \\ & (5.0 \times \text{perceived\_food\_elsewhere\_stimulus}) + \\ & (3.0 \times \text{remembered\_food\_stimulus}) \end{aligned}$$

as well as a small increment to the reservoir force if the animal's health is low. The reservoir for this system is emptied when the animal performs any one of the three consummatory actions for eating food.

### 2. avoid irrelevant animals –

$$\text{reservoir\_force} = 0.0$$

$$\begin{aligned} \text{scale\_pan\_force} = & ((260.0 \times \text{perceived\_irrelevant\_animal\_in\_same\_square\_stimulus}) \\ & +(240.0 \times \text{perceived\_irrelevant\_animal\_elsewhere\_stimulus})) \end{aligned}$$

Since there is no internal stimulus relevant to this system, there is no reservoir and there is no need to empty it.

3. **reproduce** – if there is a courted mate in the same square then

$$\text{reservoir\_force} = \text{reservoir\_force} + 0.65$$

$$\text{scale\_pan\_force} = 100.0$$

or if there is a receptive mate in the same square then

$$\text{scale\_pan\_force} = 90.0$$

or if a mate can be perceived outside the animal's square then

$$\text{scale\_pan\_force} = 50.0 \times \text{perceived\_mate\_stimulus}$$

The reservoir is emptied when the animal performs the action MATE (copulate).

4. **reduce variance** –

$$\text{reservoir\_force} = \text{reservoir\_force} + (4.0 \times \text{variance})$$

$$\text{scale\_pan\_force} = 0.0$$

The reservoir is emptied when the animal enters its den (which resets its variance to zero).

There were some problems with this initial implementation: (i) some systems, such as 'reproduce', acquired large reservoir levels when the consummatory action could not be performed for a long time. They came to dominate over everything else, even urgent needs such as to avoid predators. (ii) there was no persistence, in fact there was active dithering, since whenever a single instance of a consummatory action such as EAT\_CF was performed then the reservoir was emptied and the animal switched to another system. (iii) when external influences meant that the urgency of a sub-problem decreased of its own accord, without the occurrence of a consummatory action (e.g. when the external temperature returned towards normal and so the internal body temperature of the animal improved without any behavioural measures on the part of the animal), then the reservoir was not emptied and so the animal continued to try and do something about the problem.

## Second Implementation

These three problems were dealt with by making the following three changes to the naïve implementation: (i) ceilings were imposed on the reservoirs so that they could not increase beyond a certain maximum height. (ii) in some cases, reservoir levels were just reduced, rather than being reset to zero, when a consummatory action was performed. This is equivalent to implementing the idea of a 'sticky valve' (i.e. inertia in the movement of the valve, as proposed for Lorenz's revised model [Lorenz, 81]). (iii) reservoir levels were allowed to be reset to zero or reduced by events other than consummatory actions on the part of the animal. The following additions were therefore made to the example systems:

1. **get food** – the reservoir level was reduced by 30.0 every time the animal eats, rather than being reset to zero, and a maximum level or ceiling of 200.0 was imposed on the reservoir. When the animal was satiated with food then this condition also led to the reservoir level being reset to zero.
2. **avoid irrelevant animals** – no changes were made to this system.
3. **reproduce** – a ceiling of 200.0 was introduced for this system.
4. **reduce variance** – when the variance of the animal's estimated position decreased to less than 0.005 then this event also led to the reservoir level being reset to zero.

With this second version of Lorenz's model there were still some noticeable problems: (i) reservoir levels for systems such as 'reproduce', 'get food', and 'clean' increased throughout the night to give artificially high levels at the end of the night. (ii) it seemed impossible to tune the increases (due to internal variables and external stimuli) and decreases (due to consummatory actions and other events) in the reservoir levels so that the height of the reservoirs maintained a good correspondence with the respective deficits. After some time the deficit would be high and the reservoir level (and the



corresponding tendency for the behaviour) low, or vice-versa. This occurred partly because the ease of obtaining food, water, etc is different for different instances of the SE (with semi-random distributions of features), and also because the ease of obtaining food and water changes at different times in the life of an individual (because as it explores it finds different sources of food and water and will thereafter be able to find food and water more easily).

### **Third Implementation**

The first of the two problems mentioned above was solved by freezing all reservoir levels overnight (when the animal should be sleeping in its den). The second problem is not solvable without radically changing the whole nature of the mechanism (i.e. by making a direct link between the offset of an internal variable from its optimum value and the reservoir level, rather than the specified indirect link through the control of rate of flow through the tap) and so this was left unattended to.

The average performances ( $\bar{x}$ ) of this final, third version of the Lorenz ASM, when tested in the four versions of the SE, were 2.39, 1.40, 2.48 and 4.58, with an average over the four of **2.71**. The standard deviations ( $s$ ) of the individuals, in the tests in each of the four versions of the SE, were 2.34, 1.51, 2.38 and 4.82; and for the average result over all four versions was 2.76. The expected standard deviations of the sample averages ( $s_{\bar{x}} = (s/\sqrt{n})$ ) were 0.06, 0.04, 0.06 and 0.12; and for the average result over all four versions was 0.07.

The central idea of Lorenz's model, that unvarying levels of motivational stimuli should lead, over time, to an increasing tendency of that behaviour being expressed, is not supported by the results in this thesis. Although the phenomenon of vacuum activity seems to give evidence for this, or something similar, in the behaviour of some real animals, it does not appear to be optimal in a computational sense. The best Lorenz ASM performs less well than the best drives ASM, which is similar except in that it is



totally reactive (i.e. it only takes account of present levels of internal, indeterminate and external stimuli). The left hand side of equation 4.2 (section 4.3) is

$$\left( \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \right)$$

This is equal to

$$\left( \frac{|(6.23 - 2.71)|}{\sqrt{\frac{44.9}{6600} + \frac{10.4}{6600}}} \right) = 38.5$$

when comparing the performances of the drives and Lorenz ASMs. This is greater than 3.291 and so the probability of a real difference in the performance of the Lorenz and drives ASMs in the SE is greater than 99.9%.

Lorenz's central idea means that the tendency for a type of behaviour is dependent to a large extent on the past history of motivational stimuli for that type of behaviour. Using a similar argument to that posited by Dawkins & Carlisle [1976], when discussing parental investment, it is argued here that it is in principle sub-optimal to take any account of how long it is since the animal last engaged in this type of behaviour, or how long the animal has had this level of deficit, when calculating the desirability of engaging in behaviour that will reduce the deficit. The amount of time already invested in trying to reduce a deficit, or the amount of time that the deficit has been outstanding is not directly relevant. The only factors that are important in calculating the desirability or drive strength for a type of behaviour are those relating to the current internal and external state and future events. What has happened in the past should be of no relevance, except in as much as it allows the animal to learn how to act more efficiently in the future. An example of how the reliance on past levels of motivational stimuli can lead to anomalous reservoir levels (tendencies for behaviour) is shown in figure 9.3, in which the two levels of action specific potential should be equivalent, but are not.

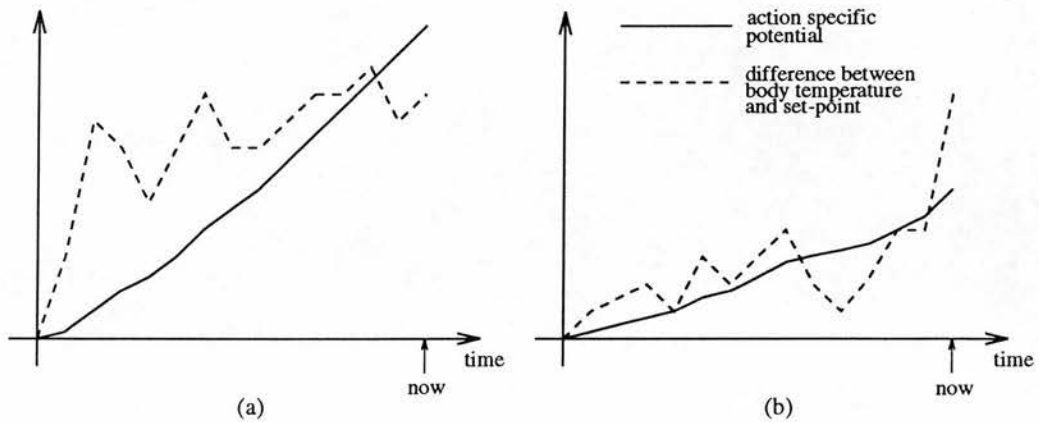


Figure 9.3: Two hypothetical graphs showing how Lorenz's model can lead to very different levels of action specific potential for the same current offset from a homeostatic set-point. In (a) the offset has been low in the recent past; in (b) the offset has been high. In neither case has there has been any consummatory action in the time interval shown.

---

The best performance for the Lorenz-type ASMs came when the reservoir levels for the homeostatic-type systems were kept roughly synchronised with the offsets (the distances from the optimal values) of the respective internal variables. But, as noted above, it proved difficult to maintain a correspondence between the two. This is hardly surprising considering that the offset of a homeostatic variable controls only the rate of inflow to the reservoir, and not the actual level itself.

In addition to the problems encountered here, somewhat similar problems have been experienced when trying to use Lorenz's model to account for observed animal behaviour:

... But to make Lorenz's model work for a wider range of behaviour, a series of increasingly complicated adumbrations become necessary. Some valves

must leak, some input taps must be connected to floats in the reservoir which progressively shut off the flow of motivation as the reservoir fills, negative pressure on the pan must be possible, and so on. Worse yet, feeding behaviour, the archetypal exemplar of the appetitive model, just does not seem to fit Lorenz's hypothesis. Motivation — hunger in this case — is controlled internally, so that overt feeding behaviour need never be performed. Something, perhaps the stretch receptors, must be imagined to be controlling another valve, releasing the accumulated motivational waters without any associated behavioural output. So, too, it is for "thirst" [Gould, 82].

The main criticism of Lorenz's model that is made here is that the size of motivational-type inputs is not tied directly enough to the current state of the environment. It should be noted that this deficit will not be specific to just this test of action selection, but will occur more generally. Also, this phenomenon is bound to occur because of a central plank of Lorenz's theory: the decision that certain stimuli should have a *cumulatively increasing* effect on the tendency for expression of behaviour, rather than a constant one. The occurrence of this phenomenon is not due to a peculiarity of this implementation of Lorenz's theory.

Although the Lorenz model of action selection did not prove particularly efficient at selecting actions appropriately, and often produced behaviour which was not particularly relevant to the current situation, there was a positive side in that the ASM did produce very coordinated, highly goal-directed activity. There was a high degree of contiguity of action sequences and the resultant behaviour seemed more 'single-minded' and purposive than that for any other type of ASM.

### 9.3.3. Performance of Maes' ASM

Maes' ASM, described in section 8.5, is a non-hierarchical, distributed network with two 'waves of input' from perceptions and motivations. There is an iteration of a spreading

activation process until one node is executable and also has an activation which is greater than the global threshold. Information about which nodes achieve or undo the different goals, which propositions about the external environment need to be true in order for a node to be relevant, and the position of nodes in consummatory/appetitive sequences, are all encoded in the different types of connections of the network.

The structure of the network that was used in the tests is shown in figure 9.4. Slightly different arrangements of connections were used on occasion, but the arrangement was always equal to, or very similar to, that in the figure. This arrangement of connections was produced because the nodes were given the precondition lists, add lists and delete lists shown in tables 9.1–9.3. For every one of the implementations many different combinations of global parameters were used to check that settings of these did not affect the observed deficits in performance. In no case was it possible to find settings of the global parameter values that removed the observed deficits.

### **First Implementation**

For the first implementation the division rules used were exactly as specified in [Maes, 89] and described in section 8.5.

Binary sensors of the environment such as 'food in square' or 'water perceived' were set to ON when the value of the perceived entity exceeded a certain threshold. Otherwise they were set to OFF. Goals were real-valued.

There was an obvious problem with this initial implementation. Because of the division by the number of outputs from the sender node of a similar type ( $N$  in figure 8.14), there was a prejudice against nodes which received input from sensors which also excited other nodes. Nodes such as 'approach perceived cover' and 'clean' received very little activation and were never chosen. Nodes such as 'drink' and 'avoid dangerous place' received a lot of activation and were chosen frequently. This principle of division by

<b>Node</b>	<b>Precondition List</b>
Eat Cereal Food	(Cereal Food in Square)
Eat Fruit Food	(Fruit Food in Square)
Pounce (Eat Prey)	(Prey in Square), (Den Not in Square)
Approach P. Food	(Perceived Food)
Approach R. Food	(Remembered Food)
Drink Water	(Water in Square)
Approach P. Water	(Perceived Water)
Approach R. Water	(Remembered Water)
Rest	(Shade or Shelter in Square)
Approach P. Shelter	(Perceived Shelter)
Approach P. Shade	(Perceived Shade)
Look towards P1	(P1 Perceived Weakly)
Freeze	(P1 or P2 Perceived), (Shelter in Square)
Run away from P1	(P1 Perceived)
Look towards P2	(P2 Perceived Weakly)
Run away from P2	(P2 Perceived)
Avoid Dangerous Place	(Dangerous Place Perceived)
Avoid Irrelevant Animal	(Irrelevant Animal Perceived)
Approach P. Mate	(Perceived Receptive Mate)
Court	(Receptive Mate in Square), (Den Not in Square)
Mate (Copulate)	(Courtied Mate in Square), (Den Not in Square)
Sleep	(Den in Square), (Sunset or Night)
Approach P. Den	(Perceived Den)
Approach R. Den	(Remembered Den)
Clean	(Den Not in Square)
Avoid Edge	(Edge Perceived)
Explore	<i>None</i>
Look Around	<i>None</i>
Approach P. Cover	(Perceived Shelter)

Node	Add List
Eat Cereal Food	(Food OK), (Health Not Too Low)
Eat Fruit Food	(Food OK), (Health Not Too Low)
Pounce (Eat Prey)	(Food OK), (Health Not Too Low)
Approach P. Food	(Cereal Food in Square), (Fruit Food in Square), (Prey in Square)
Approach R. Food	(Perceived Food)
Drink Water	(Water OK), (Health Not Too Low)
Approach P. Water	(Water in Square)
Approach R. Water	(Perceived Water)
Rest	(Not Too Hot), (Not Too Cold)
Approach P. Shelter	(Shelter in Square), (Shade or Shelter in Square)
Approach P. Shade	(Shade or Shelter in Square)
Look towards P1	(Recently Scanned), (No Predator Since Last Scan)
Freeze	(No Predator1s Perceived), (No Predator2s Perceived)
Run away from P1	(No Predator1s Perceived)
Look towards P2	(Recently Scanned), (No Predator Since Last Scan)
Run away from P2	(No Predator2s Perceived)
Avoid Dangerous Place	(No Dangerous Places Perceived)
Avoid Irrelevant Animal	(No Irrelevant Animals Perceived)
Approach P. Mate	(Receptive Mate in Square)
Court	(Court Mate in Square)
Mate (Copulate)	(Just Mated)
Sleep	(Sleeping in Den)
Approach P. Den	(Den in Square), (Variance OK)
Approach R. Den	(Perceived Shelter), (Perceived Shade), (Den in Square), (Variance OK)
Clean	(Clean), (Health Not Too Low)
Avoid Edge	(No Edges Perceived)
Explore	(Perceived Food), (Perceived Water), (Remembered Food), (Remembered Water), (Perceived Shelter), (Perceived Shade), (Perceived Receptive Mate), (Den Not in Square)
Look Around	(Recently Scanned), (No Predator Since Last Scan)
Approach P. Cover	(Close to Cover)

Table 9.2: Add list elements for each node in Maes' ASM.

<b>Node</b>	<b>Delete List</b>
Eat Cereal Food	<i>None</i>
Eat Fruit Food	<i>None</i>
Pounce (Eat Prey)	<i>None</i>
Approach P. Food	<i>None</i>
Approach R. Food	(Den in Square)
Drink Water	<i>None</i>
Approach P. Water	<i>None</i>
Approach R. Water	(Den in Square)
Rest	<i>None</i>
Approach P. Shelter	<i>None</i>
Approach P. Shade	<i>None</i>
Look towards P1	<i>None</i>
Freeze	<i>None</i>
Run away from P1	<i>None</i>
Look towards P2	<i>None</i>
Run away from P2	<i>None</i>
Avoid Dangerous Place	<i>None</i>
Avoid Irrelevant Animal	<i>None</i>
Approach P. Mate	<i>None</i>
Court	<i>None</i>
Mate (Copulate)	<i>None</i>
Sleep	<i>None</i>
Approach P. Den	(Den Not in Square)
Approach R. Den	(Den Not in Square)
Clean	<i>None</i>
Avoid Edge	<i>None</i>
Explore	(Den in Square)
Look Around	<i>None</i>
Approach P. Cover	<i>None</i>

Table 9.3: Delete list elements for each node in Maes' ASM.



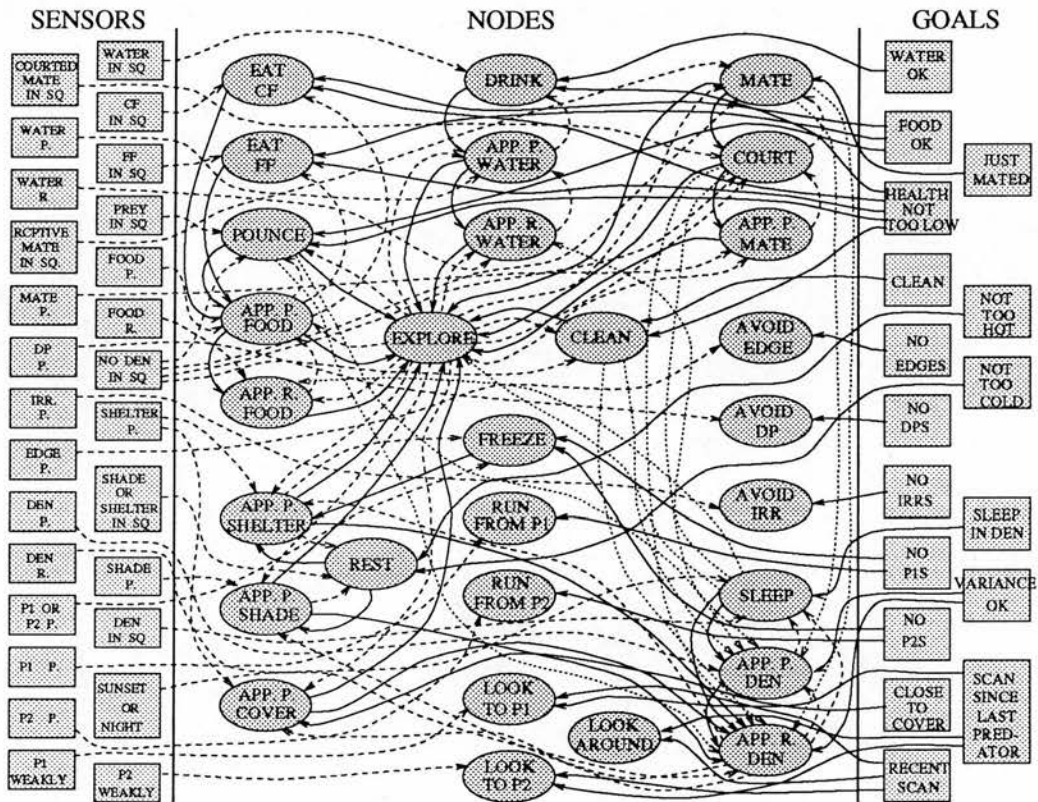


Figure 9.4: The implementation of Maes' ASM to solve the action selection problem posed by the SE. Solid lines denote goal or predecessor connections, dashed lines denote sensor or successor connections, and dotted lines denote protected goal or conflictor connections. 'P' stands for 'perceived' in sensor or node names, and 'R' for 'remembered'. 'APP' stands for 'approach' and 'SQ' for 'square'.



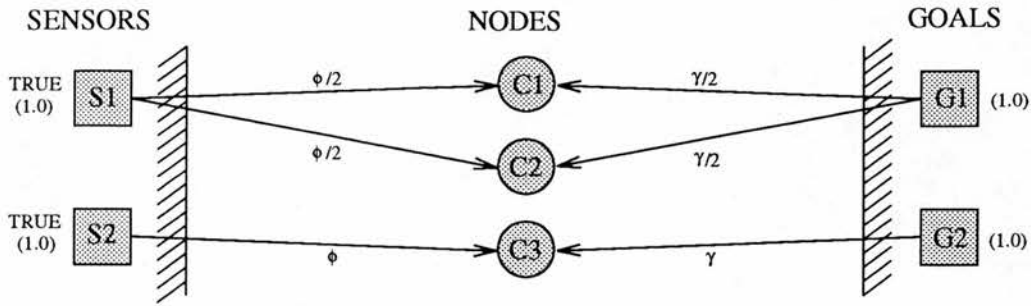


Figure 9.5: Unbalanced competition between nodes because of division by the number of outputs.  $C1$  and  $C2$  receive inputs of  $(0.5\phi + 0.5\gamma)$  whereas  $C3$  receives input of  $(\phi + \gamma)$  in every cycle, though there is no reason to prefer  $C3$ . Goal and sensor activations are shown in parentheses beside them, and the weights on the connections are shown next to the connections.

the number of outputs (or to be more exact, by the number of other nodes also receiving input from the sender node because of the proposition involved) was abandoned. This was because (i) if a goal can be achieved by several different nodes then each node should receive a full complement of excitation if it is to be able to compete effectively against other nodes (the alternative is to prejudice against nodes for which the goal that they achieve can also be achieved by one or more different nodes), as shown in figure 9.5, and (ii) if a sensor is relevant to many nodes then that should not make a difference to the amount of excitation each node receives from it (the alternative is that nodes which receive inputs from 'widely-used' sensors are penalised against).

Maes states that the division by the number of outputs is necessary because "we want modules that achieve the same goal or modules that use the same precondition to compete with one another to become active" [Maes, 89]. This argument is spurious. The modules will still compete with each other whether or not the excitation they receive is divided by the number of outputs. If the excitation from a sensor or goal is

divided up amongst the recipient nodes then these nodes will not be able to compete on equal terms against other nodes which are the only ones capable of fulfilling the goal they achieve, or on equal terms against nodes which do not share their preconditions with other nodes.

### **Second Implementation**

To create the second version of Maes' ASM the division by the number of outputs on the sender side was removed. And so, for example, the new division rules for the two cases in figure 8.14a and 8.14b were equal to  $\frac{\alpha}{M}$  and  $(\phi \times \frac{\alpha}{M})$  respectively. With these new division rules in place a second problem came to light, this time with the division by the number of inputs (or to be more exact, the division by the size of the receiving node's precondition list for environmental sensor or successor inputs, the size of the receiving node's add list for goal or predecessor inputs, or the size of the receiving node's delete list for protected goal or conflictor inputs). Because of this division there was a prejudice against nodes with many inputs of a particular type, since all of the inputs to those nodes had to be large in value in order for the nodes to accumulate as much activation as a node with only one input, which had a large value (see figure 9.6). For instance, the nodes 'look around' and 'drink' (which both receive two goal inputs, one of which is always fairly small, the other of which is sometimes quite large) are penalised because the total input they receive from goals is equal to the average rather than the sum of the inputs. This is not obviously a problem for inputs from sensors of the environment or successor links, but it is a problem for inputs from goals and inputs along predecessor links.

### **Third Implementation**

In order to remove the problem with the division by the number of goal-type inputs, and to allow nodes which can help achieve more than one goal to receive more rather than

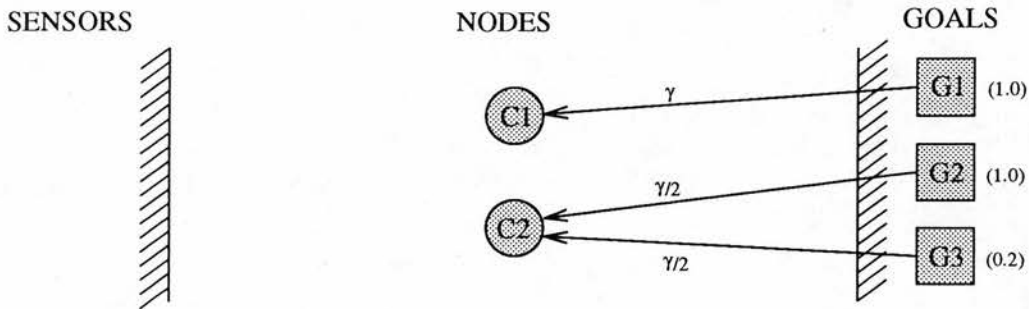


Figure 9.6: Unbalanced competition caused by division by the number of inputs.  $C1$  receives input of  $(1.0 \times \gamma)$  in every cycle,  $C2$  receives input of  $(0.6 \times \gamma)$  in every cycle, even though it can achieve goal  $G3$  as well as goal  $G2$ .

less activation, in the third implementation the division by the size of the receiving node's add list (for goals and predecessors) and by the size of the delete list (for protected goals and conflictors) was removed. This change was not made to inputs from the environment or along successor links though. And so, to give two examples, the excitations passed along links from goals were now made equal to  $(\gamma G)$ , and those along successor links to  $(\frac{\alpha \phi}{M \gamma})$ , where  $G$  is the strength of the goal,  $M$  is the number of inputs of that type,  $\alpha$  is the activation of the sending node, and  $\phi$  and  $\gamma$  are global parameters (see section 8.5).

This implementation was still problematic because nodes such as 'approach perceived food', which received many predecessor inputs, always dominated the competition. This problem was due to the change in the division rules made after the last implementation. Since there is no division by the number of predecessor inputs, nodes with many of these inputs are no longer penalised against but instead, in some cases at least, come to dominate the competition (ie. they nearly always get selected, even when they are not sensible choices).

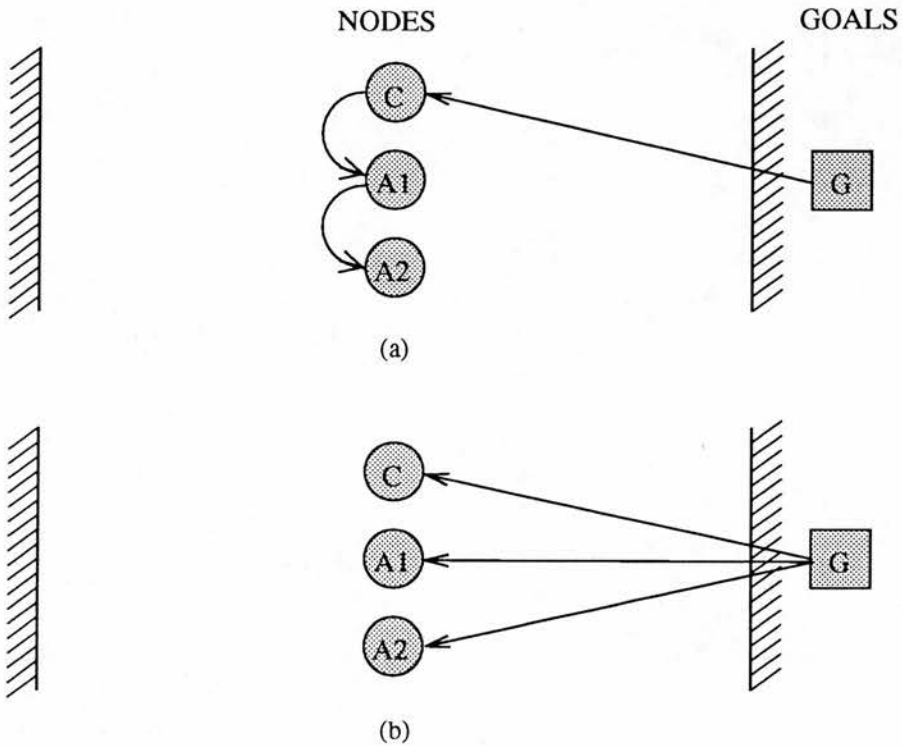


Figure 9.7: Two possible ways of propagating motivational inputs to appetitive nodes **(a)** via the consummatory node and then in a chain to successively less immediate appetitive actions, or **(b)** directly.

This problem with division by the number of predecessor or conflictor-type inputs, that both division by the number of inputs as well as the lack of such division produces unsatisfactory results, is a complex one and is due to the way in which motivational inputs are propagated through the network in Maes' ASM. Instead of a motivational variable exciting all of the appetitive and consummatory alternatives in the system (figure 9.7b and Beer's ASM in figure 8.23), in Maes' ASM the motivational variable excites only the consummatory node, and that consummatory node then passes activation back to the appetitive nodes if it isn't executable (figure 9.7a).

The difficulty with Maes' scheme for propagating motivational inputs arises when deciding whether or not to divide by the number of predecessor inputs to a node. Consider the part of the network shown in figure 9.8, in which the goal or motivational variable 'food okay' (i.e. internal food deficit) is propagated to the three nodes EAT\_CEREAL\_FOOD, EAT\_FRUIT\_FOOD and POUNCE (eat prey). When these three are unexecutable then they each pass activation through predecessor links to the node 'approach food'. In this case it is proper for activation from the three inputs to 'approach food' to be divided by the number of inputs, three, since the three eat nodes all subserve the same goal (i.e. they all act in response to the same motivational variable). On the other hand, the goals 'food okay', 'water okay' and 'just mated' (reproduce) are all propagated through consummatory and appetitive nodes to the 'furthest removed' appetitive node 'explore' (because no food, mate or water can be perceived). In this case it is not proper for activation from the different inputs to 'explore' to be divided by the number of them, since the nodes providing input to explore try to achieve different goals.

In short, as shown in figure 9.9, because the various predecessor inputs to a node can be all from nodes in the same system (trying to achieve the same goal), or can be from several systems (trying to achieve several different goals), there is no division rule for predecessor links that works for all situations. Division by numbers of predecessor inputs causes problems for nodes like 'explore', and the lack of such division causes problems for nodes like 'approach perceived food'. This difficulty with the division rules stems from the decision to feed motivational stimuli only to consummatory nodes and indirectly from there to appetitive nodes, rather than directly to both consummatory and appetitive nodes, as is done in other mechanisms (see figures 8.8, 8.9 and 8.23).

#### **Fourth Implementation**

There is no principled way of resolving this dilemma, but a tradeoff can be implemented which takes the average of the input value with division and the input value without

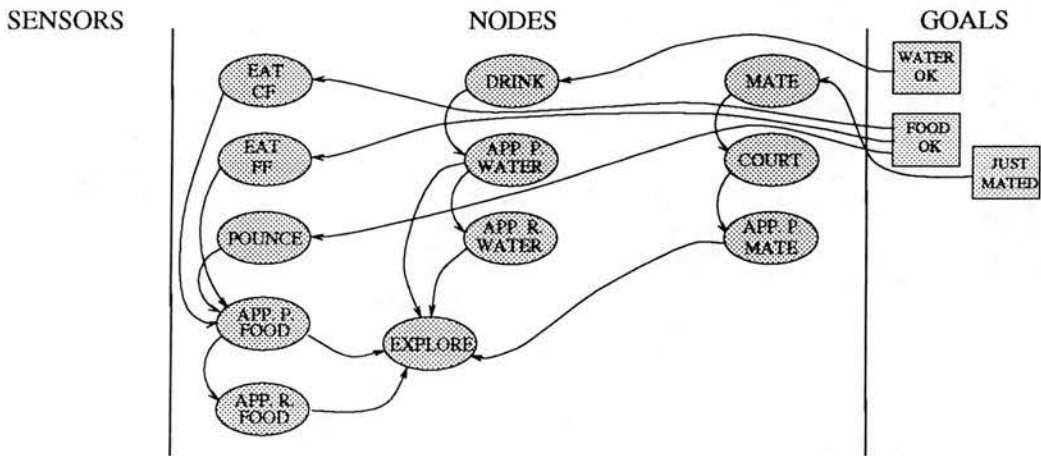


Figure 9.8: Primary goal and predecessor connections to the 'get food', 'get water' and 'reproduce' systems. 'Approach perceived food' receives three predecessor connections from nodes in the same system. 'Explore' receives five predecessor connections from nodes in three different systems.

division. So, each predecessor input is calculated according to the formula

$$I = \frac{1}{2}(\alpha + \frac{\alpha}{M})$$

and each conflictor input is calculated according to the formula

$$I = \frac{1}{2}(\alpha \frac{\delta}{\gamma} + \frac{\alpha}{M} \frac{\delta}{\gamma})$$

The above arguments do not apply to goal or protected goal inputs and so in these cases the lack of any division is maintained. It should be noted that the dilemma over division by input on predecessor links is not an attribute of the computational nature of the problem of action selection, but rather is an attribute of the way in which Maes has tried to solve it. The more usual way of passing goal information direct to appetitive nodes does not suffer from this problem.

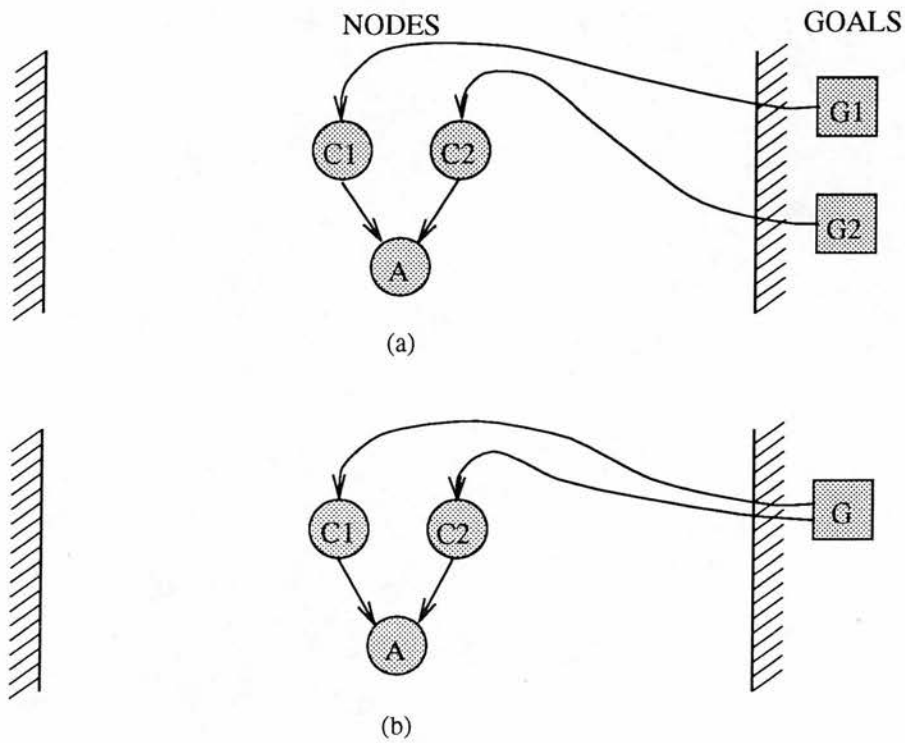


Figure 9.9: Two examples showing the dilemma when deciding whether or not to divide by the number of predecessor inputs to an appetitive node. Because there is no way of knowing whether or not the inputs derive from the same goal or not, there is no way of knowing whether there should be division by the number of inputs or not.

Using the new division rules for predecessor links, the fourth implementation was tested. The performance of the ASM was still poor, and there were problems over and above those caused by the unsatisfactory trade-off for the predecessor inputs. In particular, the main problem was that consummatory nodes seemed to get chosen only very infrequently. When the goal they could help to achieve had a high strength and when all their preconditions were true then they still lost out, on most occasions at least, to appetitive nodes in other systems. So, for instance, on occasions when 'approach water' and 'eat cereal food' were both executable (when all of their preconditions were



true), and goal strengths for 'food okay' and 'water okay' were of similar value then 'approach water' (the less optimal choice) was selected. The optimal selection is always to choose a consummatory node over an appetitive one, all other things being equal. This is because the benefits from consummatory behaviour are more immediate and more certain. Appetitive behaviour by definition does not contribute to fitness unless and until consummatory behaviour follows it.

The reason for this undesirable predominance of appetitive over consummatory nodes is illustrated by figure 9.10. In this simple example case there are two goals  $G1$  and  $G2$ , each of strength  $g$ , and each of which can be achieved only by one consummatory node ( $C1$  and  $C2$  respectively). For each consummatory node there is one appetitive node which can bring about the truth of its single precondition. If all nodes start off with zero activation then it can be shown, by calculating node activations for successive iterations of the spreading activation process, that  $A1$  will come to obtain more activation than  $C2$  after at most 5 cycles.  $A1$  will definitely obtain more activation than  $C2$  and  $A2$  after 5 cycles, whatever the values of  $G$ ,  $\gamma$  and  $\phi$ , and will overtake  $C2$  and  $A2$  after fewer cycles for certain values of  $G$ ,  $\gamma$  and  $\phi$ . This is because there is a positive feedback loop between  $C1$  and  $A1$  ( $C1$  is not executable), but not between  $C2$  and  $A2$  ( $C2$  is executable and so there is no active predecessor link).

This phenomenon of exponentially-increasing positive feedback between an unexecutable consummatory and an executable appetitive node means that executable appetitive nodes are often preferred over executable consummatory nodes in many situations when more than a few cycles are required to obtain a solution.

It was decided that it would not be sensible to continue making further alterations to Maes' ASM since this would entail a radical change in the architecture and basic concepts, with the removal or drastic alteration of the different types of connections and major changes in the way information is passed to and between nodes. When the best implementation of Maes' ASM was evaluated in the four different versions of the SE then the average results ( $\bar{x}$ ) were 0.16, 0.22, 0.18 and 0.43, with an average over the



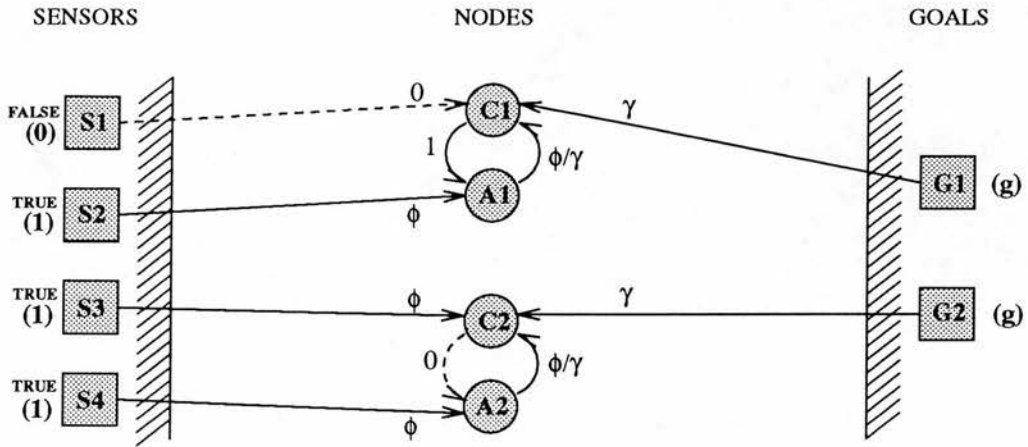


Figure 9.10: Two identical hypothetical systems, both with one consummatory and one appetitive node. Sensor and goal activations for the current state are shown in parentheses beside them, and connection weights are shown next to the connections. Dashed lines indicate inactive links because of the current state (false sensors input no activation, predecessor links only pass activation if the sender node has at least one false precondition).

four of 0.25. The standard deviations ( $s$ ) of the individuals, in the tests in each of the four versions of the SE, were 0.41, 0.49, 0.44 and 0.66; and for the average result over all four versions was 0.50. The expected standard deviations of the sample averages ( $s_{\bar{x}} = (s/\sqrt{n})$ ) were 0.01, 0.01, 0.01 and 0.02; and for the average result over all four versions was 0.01.

The left hand side of equation 4.2 is equal to

$$\left( \frac{|(2.71 - 0.25)|}{\sqrt{\frac{10.4}{6600} + \frac{0.31}{6600}}} \right) = 61.1$$

when comparing the average results of Maes model and Lorenz's model, and so the probability of a real difference in the performances in the SE is greater than 99%. Similarly, it can be shown that there is a very large probability of a real difference between the performances for Maes' ASM and the drives ASM.

There are several numerical parameters that had to be set by hand in Maes' ASM (for instance the global parameters described in section 8.5). There were also some rather arbitrary decisions that had to be made when determining the precondition list, add list and delete list for each node. A lot of experimentation was carried out with different parameter values and different lists for nodes (and also with different sets of nodes), but even so there was a danger that the problems encountered could be properties of the particular implementation of the model, rather than properties of the model itself. To counteract this, care was taken to determine the root causes of the deficits in performance, and to show that the deficits were a consequence of major decisions in the design of the model, rather than a consequence of implementational decisions made during the testing. It is argued that the major criticisms made here (that a satisfactory set of division rules does not exist, and that the mechanism does not exhibit balanced competition between consummatory and appetitive alternatives in different systems) are criticisms of phenomena that are firmly linked to a fundamental aspect of the mechanism; namely the decision to pass goal-type stimuli to appetitive nodes by way of consummatory nodes.

There are some other possible problems with Maes' ASM, which may have become apparent if not for the masking effects of the basic design faults just outlined. These are:

1. There is a loss of information because binary sensors of the external environment are assumed, whereas many properties of a realistic environment are continuous (e.g. the amount of water at a water source, the degree of protection offered by different instances of vegetation, etc).

2. Similarly, there is a loss of information because different nodes are assumed to either achieve a goal or do not. There are no ways of expressing the fact that the action/behaviour generated by a node can achieve a goal to some degree, or with some probability. This is also the case with negating goals — Maes assumes that nodes either counteract a goal or do not. There is no way of encoding the fact that, for instance, strenuous actions such as MATE tend to increase body temperature (i.e. help to achieve the goal 'keep warm' to some degree) and also tend to increase the rate at which food and water are used up (and hence help to partially undo the goals 'food okay' and 'water okay').
3. Persistence – the necessary mechanisms to ensure persistence are not all present in this ASM. Although predecessor and successor connections tend to promote contiguous sequences of behaviour (e.g. explore for food, approach food, then eat it), there is nothing to favour continued eating at a food source after the animal has already eaten there once. In fact, the animal is less likely to continue eating in the next timestep since the internal deficit will have decreased and the activation of the winning node is reset to zero, whereas all other nodes retain some of the activation they finished with in the previous timestep.
4. The decision to add activation when preconditions are true unfairly penalises those nodes with no preconditions (such as 'explore' or 'look around'). Tinbergen made a distinction in his mechanism between stimuli which increase the activation of the node, and stimuli which impinge upon the innate releasing mechanism (thus tending to 'unblock' the node, but not to make it more likely to get selected).
5. In line with Lorenz, Maes assumes that excitation from stimuli should always be summed when calculating node activation levels. This is not always the case in real animals. Some animals seem to respond in proportion to the product, rather than the sum, of two different stimuli. This is discussed further in section 10.4.

In summary, the lack of suitable division rules and the unbalanced competition between consummatory and appetitive alternatives combine to produce some serious deficits in

the performance of this ASM. These problems were discussed in detail and shown to be problems that are not specific just to this SE. They will occur in general whenever there are multiple goals, consummatory and appetitive alternatives, and many candidate actions to choose among. This mechanism has been reported by Maes [1989a] to work well at planning-type problems, but it does not seem appropriate for animal-like action selection problems such as that posed by the SE.

### 9.3.4. Performance of Rosenblatt & Payton's ASM

#### First Implementation

This ASM, a hierarchy with expression of preferences and unrestricted flow of information, was described in section 8.6. It is strictly feed-forward with no recurrent connections and no inhibitory or excitatory connections to other nodes at the same level. The implementations of this ASM (especially the third) are described rather more briefly than those for other ASMs because this ASM is covered in more detail in chapter 11.

The function which calculates each node activation from the preferences for it is unspecified by Rosenblatt & Payton, [1989] (see figure 8.15). For this first implementation it was constrained to be a simple sum of the preferences for the node. Exceptions to this occurred with *cue stimuli*, i.e. stimuli which signal the appropriateness or otherwise of an action (e.g. EAT.FRUIT.FOOD is only appropriate if the stimulus 'fruit food in square' is greater than zero). In these cases the preference is multiplied with the cue stimulus input before being combined with other preferences so that there will be no activation passed to the node unless it is both relevant to the current situation and there is a need for it.

Four example systems from the first implementation are now described.

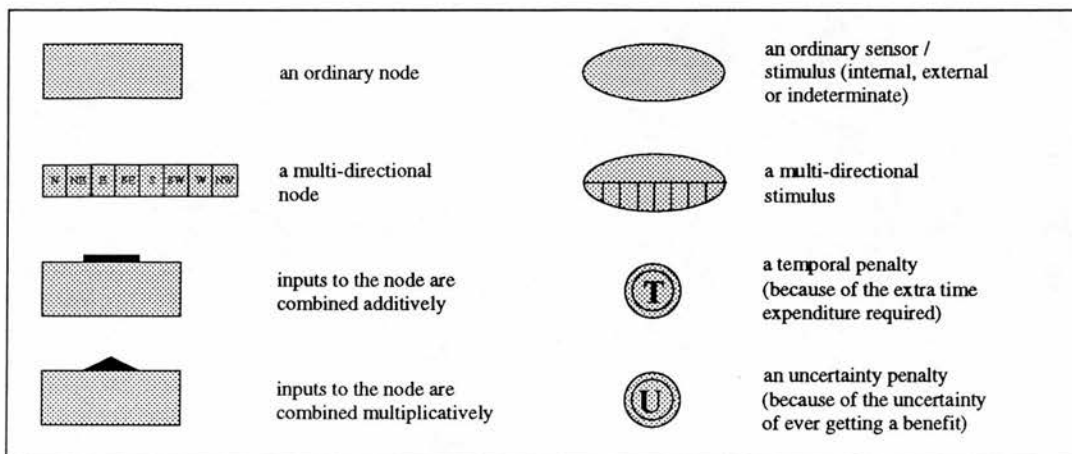


Figure 9.11: Key to Rosenblatt & Payton ASM figures. When a node is shown with neither a shaded triangle nor a shaded rectangle above it then the node has a more complicated rule for combining preferences.

Figure 9.12 shows the 'get food' system. The internal stimuli 'food minus' and 'low health' impinge on the top-level node 'get food'. In this simple implementation the four alternative sub-systems 'eat food', 'approach perceived food', 'approach remembered food' and 'explore for food' are all activated equally (there is also a small amount of excitation to 'don't use up food'). For each of the three eat actions, there is a separate cue-type stimulus, which is multiplied with a propagated motivational stimulus so that the eat actions only receive activation if the relevant type of food is present. So, for instance, the activation passed to EAT\_FRUIT\_FOOD equals

$$(\text{excitation from 'eat food'}) \times (\text{stimulus from 'fruit food in same square'})$$

'Perceived food', 'remembered food' and 'random direction' are all stimuli with eight different components, one for each direction the animal can move in. Each component of each stimulus is multiplied with the relevant motivational-type input (i.e. an input from a node that is higher up in the hierarchy). To give an example, the activation of 'move north' is equal to

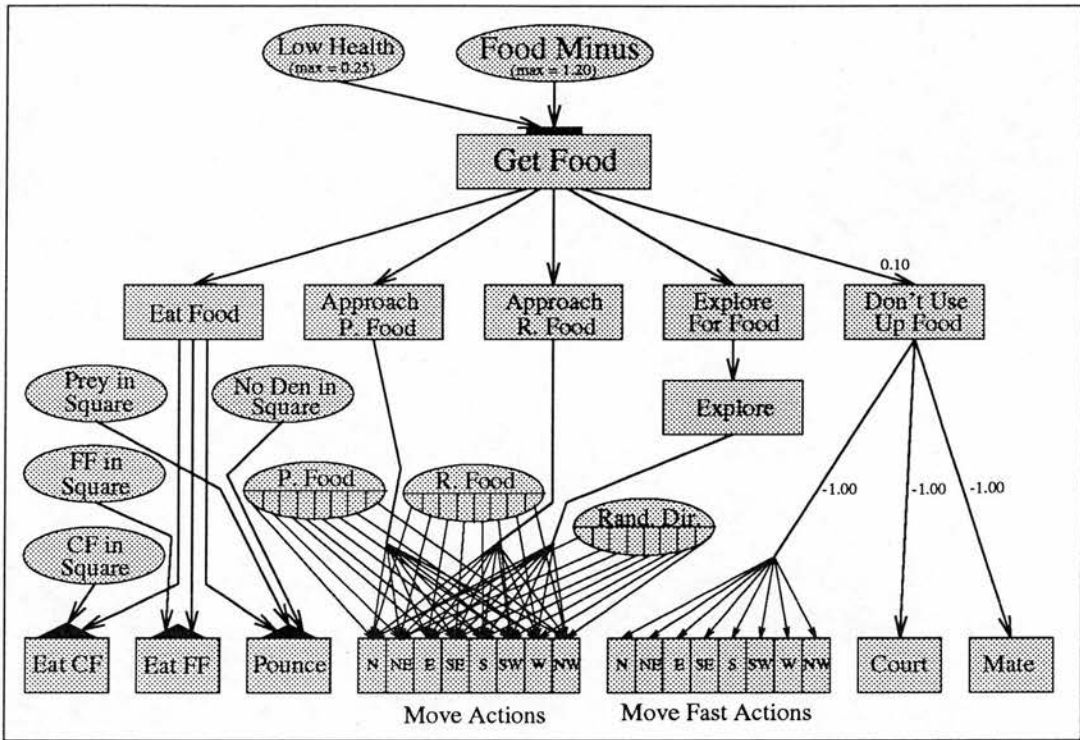


Figure 9.12: 'get food' system from the first implementation of Rosenblatt's & Payton's mechanism. The numbers next to connections are weights. Connections from other systems are not shown. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

$$[(E_1 \times S_1) + (E_2 \times S_2) + (E_3 \times S_3) + \text{inputs from other systems}]$$

where  $E_1$  = the excitation from 'approach perceived food',  $S_1$  = the stimulus value of 'fruit food to north',  $E_2$  = the excitation from 'approach remembered food',  $S_2$  = the stimulus value of 'remembered food to north',  $E_3$  = the excitation from 'explore', and  $S_3$  = the stimulus value of 'random direction to north'. The  $(E \times S)$  pairs are summed because that is the rule for combination of preferences in this first implementation.

An example of the flow of activation in this system is shown in figure 9.13.



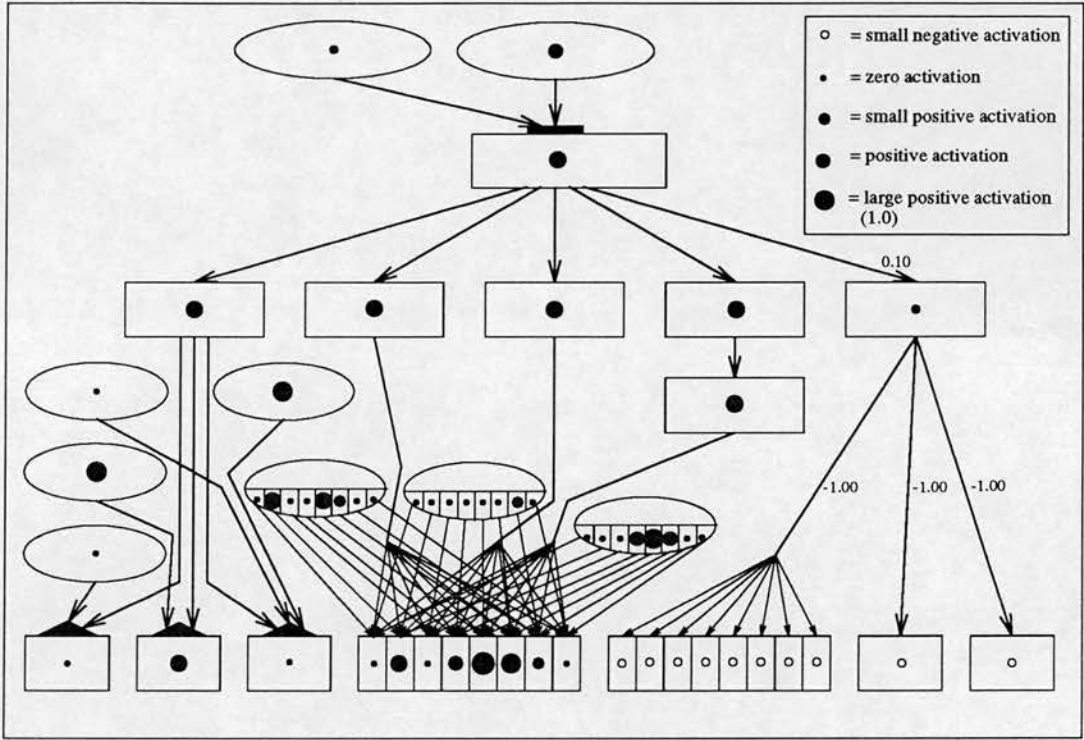


Figure 9.13: An example of the flow of activation through the 'get food' system. The effects of other systems (e.g. on the MOVE nodes) are not shown. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

Figure 9.14 shows the 'keep clean' system. The system has two inputs from 'low health' and 'dirtiness' (i.e.  $1.0 -$  the value of the internal variable 'cleanliness'). The animal can not clean itself in its den, and so there is a need for an appetitive action to leave its den if necessary. The multi-directional stimulus 'all directions' is equal to 1.0 in all directions.

Figure 9.15 shows the 'avoid predator1s' system. There are five different alternative sub-systems, although these cannot be categorised as consummatory or appetitive (except for 'approach shelter' and 'freeze in shelter'). There is a large, constant, motivational input of 3.9. Again, though, this will not affect the activations of any action-level

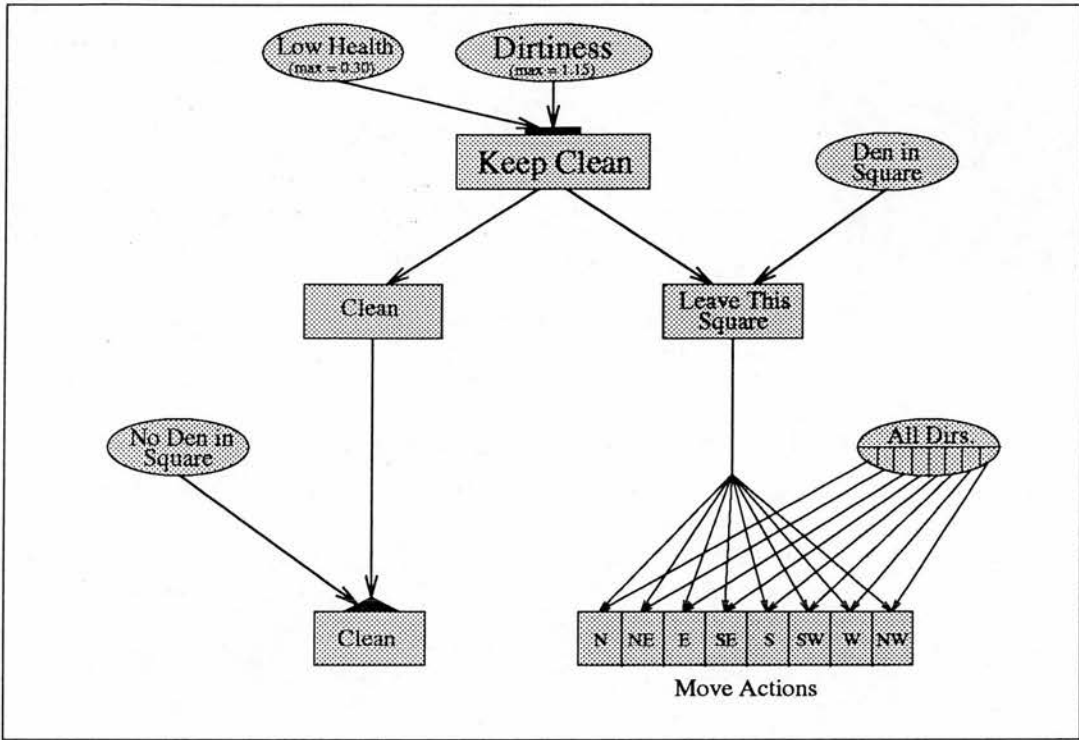


Figure 9.14: 'keep clean' system from the first implementation of Rosenblatt's & Payton's mechanism. Connections from other systems are not shown. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

nodes unless a predator1 is perceived. 'Max P1' is the maximum value, over the 8 different directions, of the perceptions of predator1s. 'Max P1 (adj)' is 0.0 if 'Max P1' is less than 0.30, otherwise it is equal to 'Max P1'. This is required to stop the animal choosing the action FREEZE too frequently when it is in dense vegetation (which produces a large value of 'shelter in square' and also reduces the efficacy of perception, producing many incorrect, low-valued perceptions of predator1s). 'Opp P1' is zero in the direction of the predator1s and adjacent directions, 1.0 in the exact opposite direction and then slightly less than 1.0 in the directions adjacent to the opposite one.

The final example system (figure 9.16) is the 'vigilance', or 'detect predators' system.



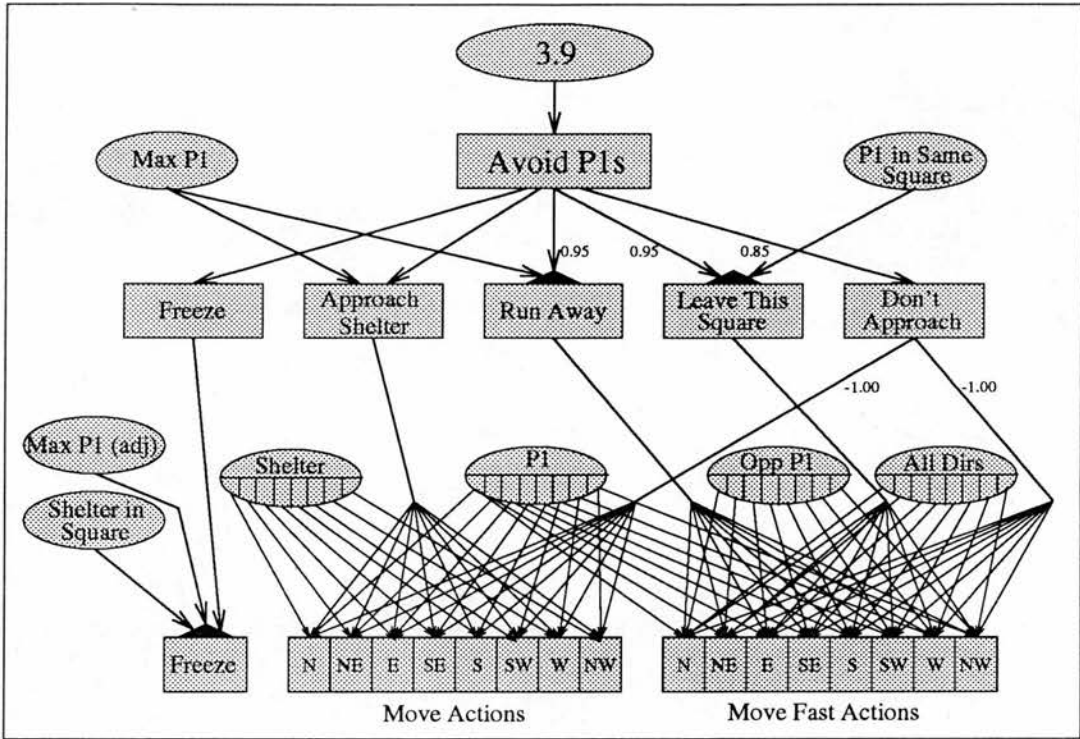


Figure 9.15: 'avoid predator1s' system from the first implementation of Rosenblatt's & Payton's mechanism. Connections from other systems are not shown. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

This has two motivational inputs to the system-level node, one proportional to the time since the last look-type action, and the other which is high if a predator has been perceived since the last look-type action was performed. The action LOOK\_AROUND is not dependent on any external stimulus. It will always lose out to an action to look directly at a predator, if a predator is perceived. The two stimuli 'P1 (adj)' and 'P2 (adj)' are transformed multi-directional perceptual inputs which have been altered so that they are equal to 1.0 if the same component of 'P1' is greater than 0.2, and are multiplied by  $\frac{1}{0.2}$  otherwise. This is because the animal needs to look in the direction of uncertain perceptions of predators.

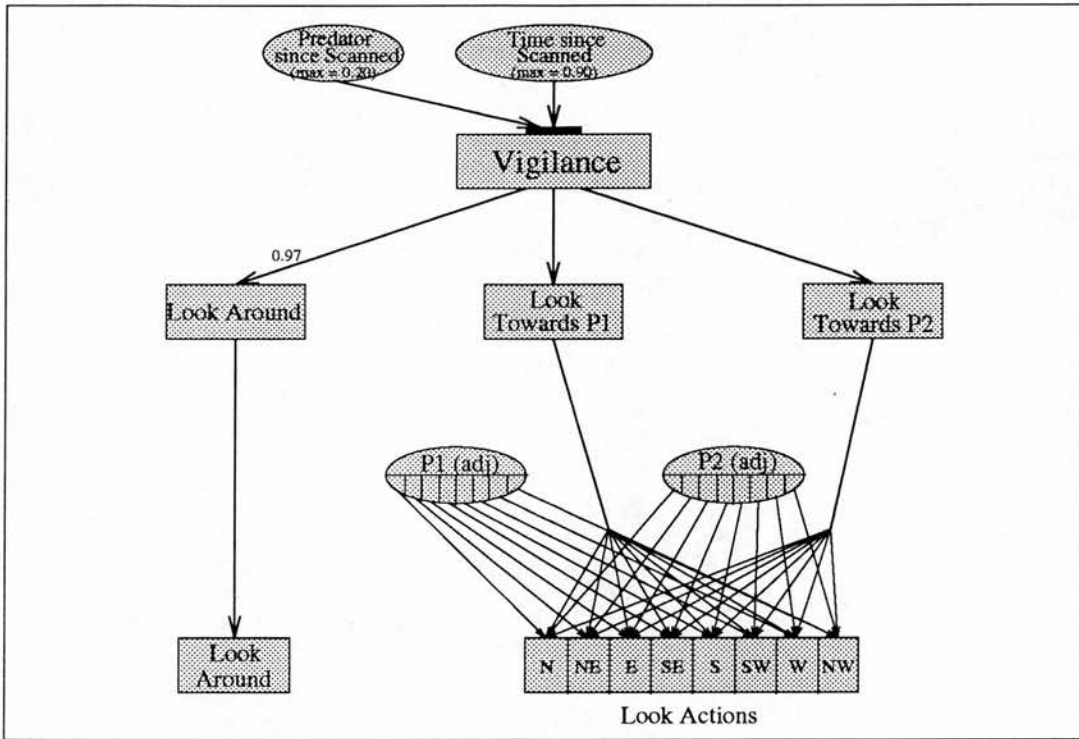


Figure 9.16: 'vigilance' or 'scan for predators' system from the first implementation of Rosenblatt's & Payton's mechanism. Connections from other systems are not shown. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

This first, naïve implementation of the Rosenblatt & Payton mechanism was tested in the SE and one particular problem was found. The MOVE actions, which receive more preferences than any other action nodes, always win the competition, and so the animal spends all of its time moving in one direction or another and actions such as EAT\_FRUIT\_FOOD or DRINK or MATE never get chosen.

One reason for this is that preferences from appetitive sub-systems (which are usually to MOVE actions) are similar in size to those from consummatory sub-systems, which ignores the fact that consummatory actions are obviously more preferable, since they

lead to achievement of the goal (which should imply an increase in future expected genetic fitness) more immediately, and more certainly. In order to rectify this problem, Sutton & Barto's idea of *temporal discount factors* (see section 8.8.5) was used in an amended form to yield a second implementation.

## Second Implementation

In addition to penalising those options which achieve their reward over a longer time-scale, an additional penalty was introduced here for those options which were less certain to lead to an eventual reward. Therefore two types of inhibitory inputs were introduced: (i) *temporal penalties* which inhibit in proportion to the length of time to likely achievement of the goal (e.g. 'explore for food' might have a 0.5 temporal penalty because it is not likely to lead to actual ingestion of food until after a fairly long period of time), and (ii) *uncertainty penalties* which are inversely related to the likelihoods of the options eventually contributing to achievement of the goal. For example, 'approach remembered water' might have a 0.3 uncertainty penalty because there is a fairly high probability that even if the animal decides to head towards a remembered water source it will be interrupted en route, for instance by an encounter with a predator or mate. In that case it would probably not continue to approach the remembered water source and so the original action was wasted.

These two types of penalties are largest for those appetitive nodes which are furthest removed from the consummatory node. Further discussion of the different means that could have been used to alter the mechanism to favour consummatory over appetitive options, and why this particular method was chosen, is given in section 11.1.4.

To give an example of the penalties used in two systems, the different appetitive options in the 'get food' system had temporal and uncertainty penalties as shown in figure 9.17a, and in the 'avoid predators' system the penalties were as shown in figure 9.17b. The other two example systems, 'avoid hazards' and 'detect predators' do not have

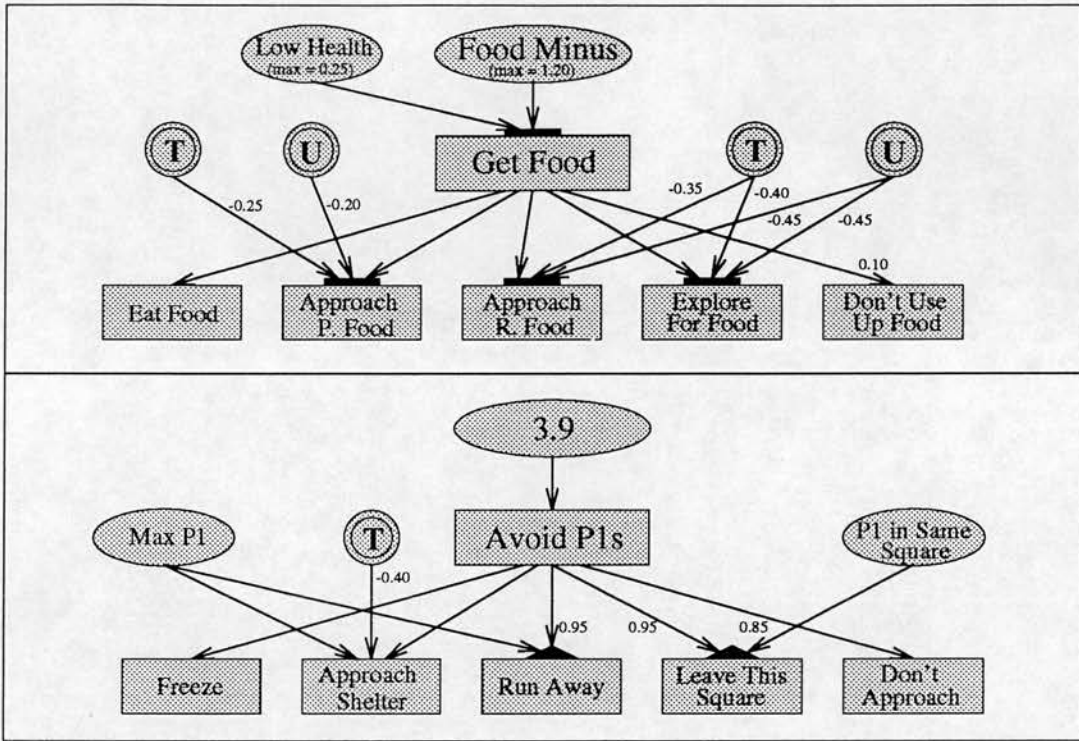


Figure 9.17: Temporal (T) and uncertainty (U) penalties applied to appetitive nodes in the 'get food' and 'avoid predator1s' systems in the second implementation of the Rosenblatt & Payton mechanism. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

consummatory and appetitive options and consequently there were no temporal or uncertainty penalties. With the penalties added the Rosenblatt & Payton mechanism was tested a second time. This time the MOVE actions were made less dominant by imposing severe penalties on the appetitive options. However, the problem did not entirely go away because there were still several unaffected inputs to the move actions, from non-appetitive sub-system nodes. Another problem that emerged was that, because of the penalties on them, appetitive options were not able to accrue enough activation on their own accord. So, for instance, even if the animal was very

short of food then it would not approach a clearly perceived food source because of the two severe penalties on 'approach perceived food'. The animal needs to prefer an eating action over an approach food action if the two have similar stimulus values, but at the same time still needs to attach a fairly high importance to appetitive options like 'approach perceived food' or 'explore for food' if the animal is very short of food and there is no food immediately at hand, or none perceived or remembered sufficiently strongly.

### Third Implementation

The solution to these two problems was to make the penalties on appetitive options less severe (e.g. figure 9.18) and at the same time to change the combination of inputs rule from a simple sum to that shown below

$$A_j = \left( \frac{\max_i(P_{ij}^+) + \alpha \sum_{i=1}^{N^+}(P_{ij}^+)}{1 + \alpha} \right) + \left( \frac{\min_i(P_{ij}^-) + \beta \sum_{i=1}^{N^-}(P_{ij}^-)}{1 + \beta} \right)$$

where  $P_{ij}^+$  and  $P_{ij}^-$  are the positive and negative preferences from node  $i$  for node  $j$ ,  $N^+$  and  $N^-$  are the numbers of such preferences for node  $j$ , and  $\alpha$  and  $\beta$  are constants. With fairly small values of  $\alpha$  and  $\beta$  this prevented the MOVE actions from being selected on every occasion, and also allowed appetitive sub-systems to exert a large influence on the chosen action when appropriate.

The computational reasons why a simple sum is not satisfactory are explained more fully in section 11.1.3. Other parts of chapter 11 also contain detailed consideration of other aspects of this ASM.

With the more moderate penalties and the new rule for combining preferences, this third implementation performed very competently, with average results ( $\bar{x}$ ) in the four different versions of the SE of 8.09, 3.61, 8.16 and 13.38, with an average over the four of 8.31. The standard deviations ( $s$ ) of the individuals, in the tests in each of the four versions of the SE, were 7.06, 3.63, 6.66 and 9.16; and for the average result over

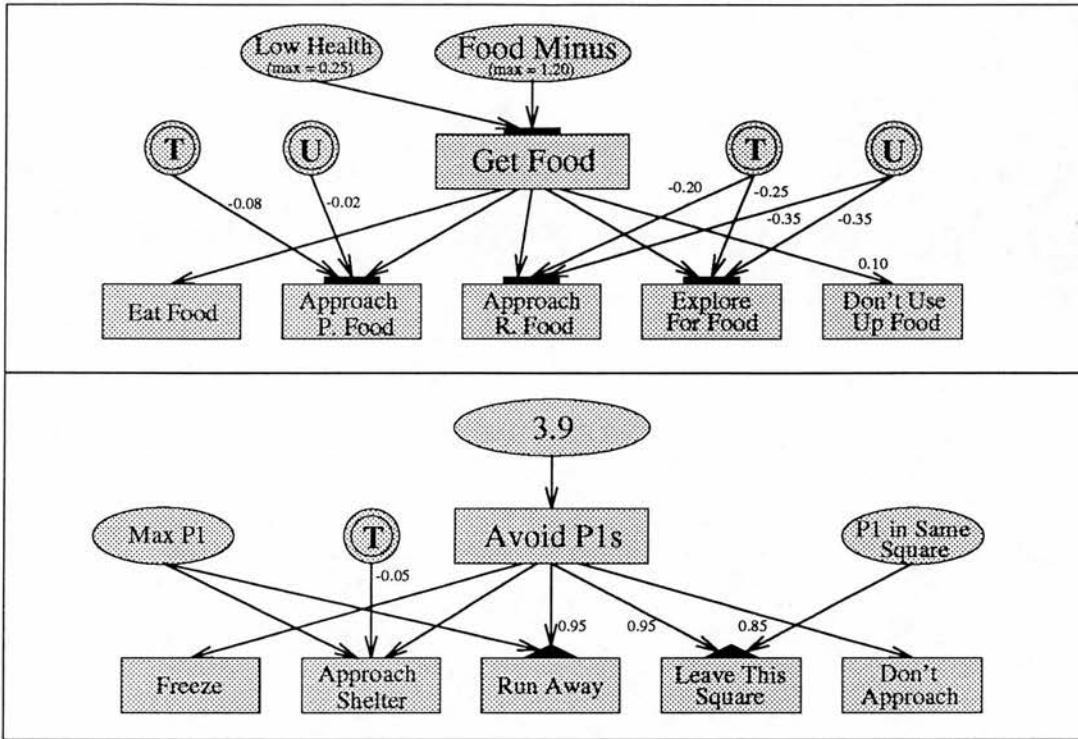


Figure 9.18: Less severe temporal and uncertainty penalties applied to appetitive nodes in the 'get food' and 'avoid predator1s' systems in the third implementation of the Rosenblatt & Payton mechanism. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

all four versions was 6.63. The expected standard deviations of the sample averages ( $s_{\bar{x}} = (s/\sqrt{n})$ ) were 0.17, 0.09, 0.16 and 0.23; and for the average result over all four versions was 0.16.

It can be seen that a modified Rosenblatt & Payton mechanism (including the improvements to the original design) is able to perform more competently than any of the other ASMs tested here. Equation 4.2 (section 4.3) can be used to show that there is a greater than 99% probability of a real difference between the performance of this ASM and the performances of the drives ASM (the next best one), since

$$\left( \frac{|(8.31 - 6.23)|}{\sqrt{\frac{69.6}{6600} + \frac{44.9}{6600}}} \right) = 15.8$$

which is greater than 3.291. Equation 4.2 can similarly be used to state with a high degree of certainty that the modified Rosenblatt & Payton mechanism performs better than the Lorenz and Maes ASMs in the SE.



## Chapter 10

# Theoretical Analysis

Some theory was touched upon in the last chapter in the course of unravelling the causes of certain deficits in the different ASMs. For example, the need for persistence, the need to prefer consummatory over appetitive alternatives, and problems with passing motivational stimuli to appetitive nodes indirectly through the consummatory nodes were all mentioned.

The first four sections of this chapter present some additional theoretical points which, together with those in the previous chapter, provide a framework for the analysis and criticism of the different mechanisms. This theory will be useful for explaining more subtle differences between the performances of different ASMs. Following on from these theoretical discussions, the fifth section of this chapter consists of a list of considerations that should be taken into account in the design of any future ASMs. The sixth section enumerates the shortcomings of each of the ASMs that were tested in the SE, and the final section summarises all of the testing results.

### 10.1 Dealing with All Types of Sub-Problem

The variation in possible types of sub-problem was discussed in section 7.2, and a terminology to describe them was proposed. It is important that, for instance, ASMs are



not designed solely to deal with homeostatic, internal stimulus dependent, prescriptive sub-problems such as 'get food' (the most common sub-problem considered in the ethological literature). Lorenz's model of action selection or behavioural choice is oriented towards these sorts of sub-problems, and is rather inappropriate for sub-problems such as 'avoid hazards'. A satisfactory mechanism needs to be able to handle all different types of sub-problem appropriately.

## 10.2 Compromise Candidates

One of the advantages of Rosenblatt & Payton's mechanism (see section 8.6) is that their mechanism is able to consider preferences of more than one system simultaneously. Most ASMs allow only one system to be active (and express preferences) at any one time, whether through intra-layer inhibition or some other means (see section 11.1.4). All but one of the systems is 'shut down' and no longer takes part in the selection process. Rosenblatt & Payton's idea of combining evidence from all systems when choosing an action allows, in some cases, better choices to be made.

For instance, proscriptive sub-problems such as 'avoiding hazards' should place a demand on the animal's actions that it *does not* approach the hazard, rather than positively prescribing any particular action. It is obviously preferable to combine this demand with a preference to head towards food, if the two don't clash, rather than to head diametrically away from the hazard because the only system being considered is that of 'avoid hazard' (see figure 10.1).

More generally, a *compromise candidate*, which might be beneficial to two or more systems to an intermediate degree, may be preferable to any of the candidates which are most beneficial for one system alone.

Although computational constraints, described in section 11.1.3, mean that pure combination of preferences is not always straightforward, it should still be incorporated in modified form because of the ability to choose compromise candidates that it provides.

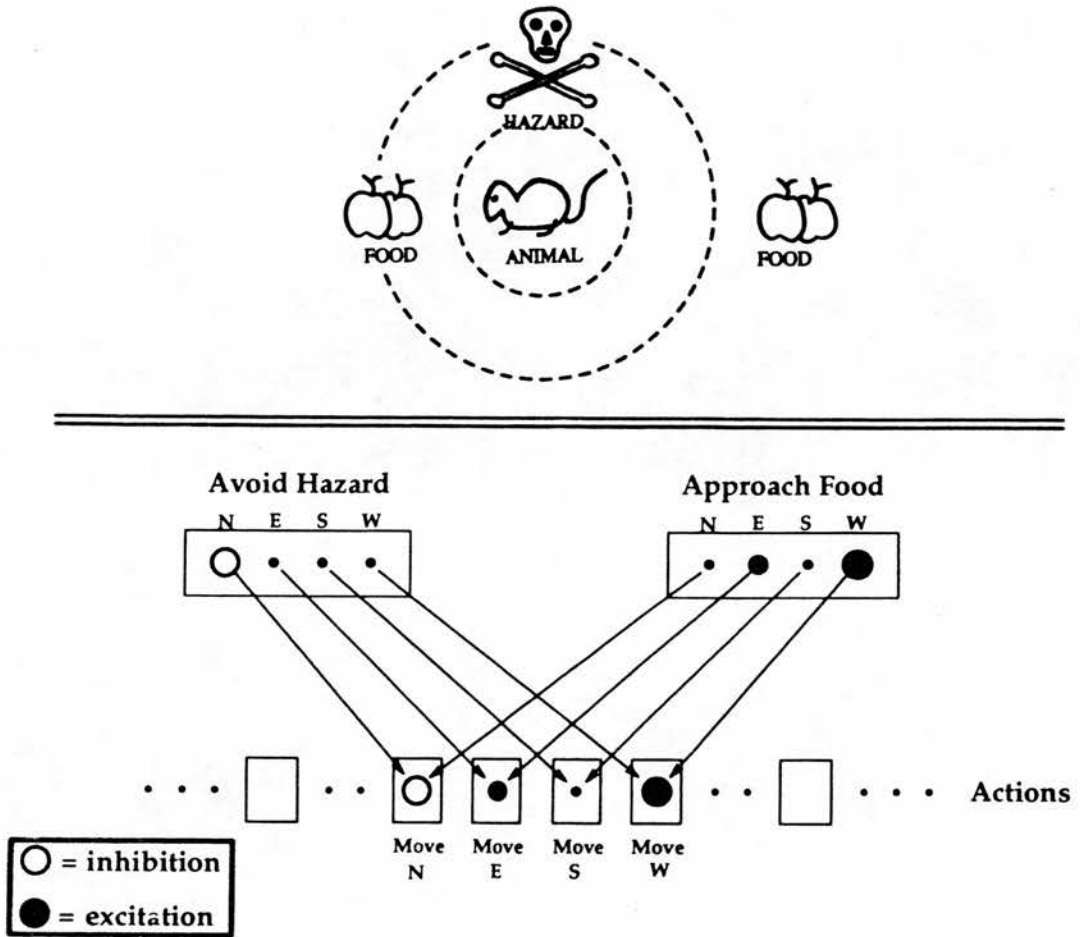


Figure 10.1: Choosing a compromise candidate by combining preferences from proscriptive and prescriptive systems (chosen action is MOVE\_SOUTH if the needs of only one system can be considered at one time, and is MOVE\_WEST if combination of preferences is allowed). Larger circles denote greater levels of excitation or inhibition.

### 10.3 Proper Treatment of Information

Many researchers into the workings of the brain have stressed the importance of considering the treatment of information. Neural networks today are often analysed partly in terms of whether they manipulate the information provided as input to them in a principled manner. The main aspect of the treatment of information that is of importance here is that of *preservation of information*. In order to be able to make

optimal decisions, an ASM must not throw away any relevant information. This need to preserve relevant information leads to two considerations:

1. **characterising the internal and external state** — some ASMs (e.g. Maes' one) include binary-valued sensors of the environment. These virtual sensors return 1 or 0 according to the truth or falsity of logical propositions about the environment (such as 'perceived food stimulus  $> 0.3$ '). The use of such binary or integer sensors results in a loss of relevant information because the mechanism cannot distinguish between, for example, a food stimulus of value 0.35 and a food stimulus of value 0.95. As a general rule, sensors of internal and indeterminate variables and the external environment should be real-valued.
2. **maintaining preferences of all systems** — many of the mechanisms considered here select one system as the most appropriate and *then* select one action based only on the needs of that system. As discussed above, this is sub-optimal because it does not allow for combination of preferences and selection of compromise candidates. Looked at from another point of view, selecting just one system is not a good idea because it leads to a *loss of potentially useful information* (i.e. the loss of the preferences of all of the 'switched off' systems). The principle of preservation of information requires that the preferences of all systems are propagated through to actions until a decision has to be made at the level of the behavioural final common path.

## 10.4 Combination of Stimuli

How should the various stimuli relevant to a system, sub-system or action be combined? Should the stimulus magnitudes be added in order to calculate the activation or 'degree of preference' for that node? Or should they be multiplied? Arguments put forward in [McFarland, 85] suggest that whereas some stimuli combine in an additive fashion (figure 10.2a), others combine in a multiplicative fashion (figure 10.2b). But it is possible, and perhaps likely, that more complex functions are sensible in certain

situations. Consider, for example, the calculation of the 'get food' drive strength in the drives ASM. The degree of preference, or degree of benefit to the animal, of choosing the system 'get food' depends on both internal stimuli (food deficit) and external stimuli (perception of food). It would seem that both simple additive and simple multiplicative rules are inadequate in this case. The simple additive rule

$$P_{get\_food} = IFD + FS$$

where  $P_{get\_food}$  = preference for system 'get food',  $IFD$  = internal food deficit and  $FS$  = perceived food stimulus, is unsatisfactory because it can be greater than zero when the animal is satiated with food ( $IFD = zero$ ). The simple multiplicative combination rule

$$P_{get\_food} = IFD \times FS$$

is also unsatisfactory because the product is equal to zero if  $FS = zero$ , even if  $IFD$  is rather high. If the animal is nearly starving then it should still try and get food (e.g. by exploring) even if it cannot perceive food at the current moment.

The exact combination of stimuli that should be employed in this case is not the issue. What is important here is that an ASM should be able to accommodate different rules for combination of stimuli, and should not presuppose strict summation or multiplication.

## 10.5 Requirements for an Action Selection Mechanism

Several theoretical issues relevant to action selection / behavioural choice have now been discussed in chapter 9 and the previous sections of this chapter. This section now lists specific points that should be taken into account in the design of future ASMs. These points can alternatively be thought of as a set of benchmark criteria, each of which is failed by one or more of the ASMs tested here. These criteria can also be used to give an indication of the suitability of future proposed ASMs.

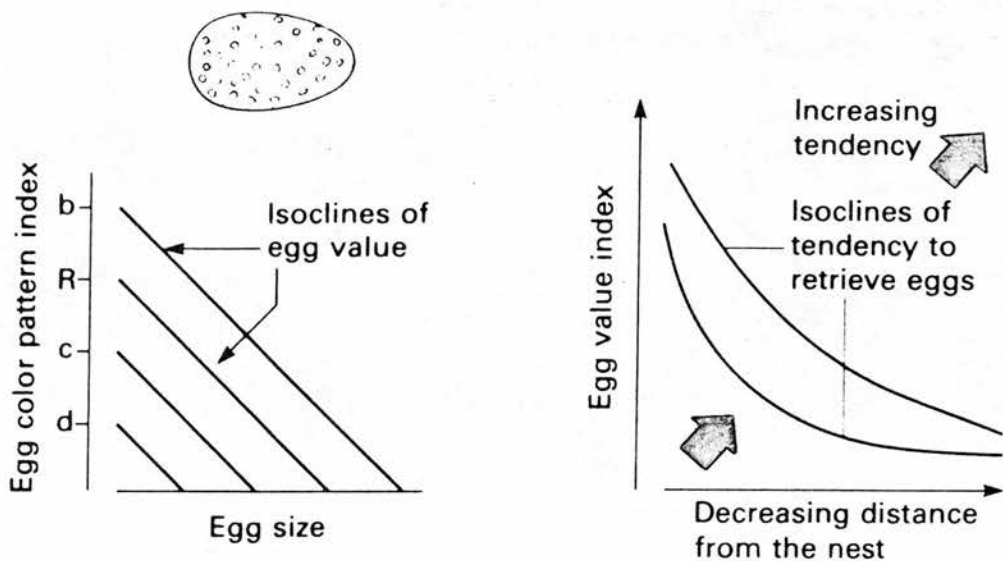


Figure 10.2: (a) additive effects of two stimuli, (b) multiplicative effects of two stimuli. Taken from [McFarland, 85].

1. **Dealing with all types of sub-problem:** the need to be able to handle all of the various types of sub-problem successfully (see sections 7.2 and 10.1).
2. **Persistence:** the need to have a tendency to persist with a consummatory action beyond the time that the deficit it is reducing is strictly the most important, because of the 'cost' of changing to another system. The cost is the amount of time it will take to obtain a situation in which the consummatory action for that other system can be performed, as described in section 9.3.1.
3. **Activations proportional to current offsets:** in homeostatic systems, the need for the node activations or drive strengths to be in proportion to the *current* offsets from the optimal set-points (see section 9.3.2).
4. **Consummatory over appetitive actions in the same system:** the need to have a tendency to prefer consummatory over appetitive actions in the same system, if both are equally relevant to the current external situation (see sections 9.3.3 and 9.3.4)

5. **Consummatory over appetitive actions in the other systems:** the need to have a tendency to prefer a consummatory action in one system over appetitive actions in other systems, all other things being equal (see sections 9.3.3 and 9.3.4)
6. **Balanced competition:** the need for there to be no discrimination against nodes which help to achieve more than one goal, or against nodes which receive input from many different stimuli, or against nodes which receive input from only one stimulus or from none at all (see section 9.3.3 and 9.3.4).
7. **Contiguous action sequences:** for a similar reason to that underlying the need for persistence (i.e. the cost of changing), there is a need to have a tendency towards continuing the current sequence once started, rather than beginning a new sequence for a different system (see section 9.3.2).
8. **Interrupts if necessary:** the need to be able to interrupt a sequence of actions for a relatively low-priority system (e.g. 'get food') if another more urgent system (e.g. 'avoid predators') places a high-priority demand on the use of the animal's actions (see section 9.3.2).
9. **Opportunism:** the need to incorporate external stimulus ('availability') information, as well as deficit or motivational information, when calculating the desirability of different alternatives (see section 9.2.5). This should allow the animal to interrupt other activities to take advantage of infrequently-available opportunities (for instance, a receptive mate in the same square) if they should suddenly arise.
10. **No system-level winner-take-all:** because of the needs for preservation of information and choice of compromise candidates, a mechanism should *not* 'shut down' all but one system (see sections 10.2 and 10.3).
11. **Combination of preferences:** the need to be able to integrate multiple non-binary preferences from higher-level nodes when deciding the responses of lower-level nodes (see sections 10.2 and 10.3).

12. **Compromise candidates:** the need to be able to choose actions that, while not the best choice for any one sub-problem alone, are best when all sub-problems are considered simultaneously (see sections 10.2 and 10.3).
13. **Real-valued sensors:** the need to extract the full amount of information from the environment and internal state (see section 10.3).
14. **Flexible combination of stimuli:** the need to allow arbitrary functions for combining stimulus values (see section 10.4).

## 10.6 Summary of Findings on Each ASM

This section summarises the positive and negative aspects of each of the four ASMs tested in the SE.

Two potential concerns over these results are that *(i)* the deficits are not applicable in general, but are specific to this test of action selection, and *(ii)* the deficits would not occur if the ASMs had been implemented more competently (the deficits could be a property of this particular implementation of the ASM rather than of the ASM as a whole). To guard against these concerns, care was taken in each case to ensure that: *(i)* the *reasons* behind each deficit in performance were determined, rather than just accepting them at face value, *(ii)* the reasons were generally applicable to problems of action selection, rather just this one, *(iii)* a lot of time was spent experimenting with different equations, parameter values, preconditions, etc., and *(iv)* the reasons or causes underlying the deficits were due to central features of the ASM, rather than minor details which could be changed easily. These safeguards, taken together, ensure that the deficits that are described here are applicable to action selection in general, and arise as an inevitable consequence of central tenets of the ASM in question.



### 10.6.1. Drives

The drives model is the simplest considered here, partly because it only addresses half of the problem, i.e. how to select a system. The naïve initial implementation had no means of incorporating opportunism and, as predicted, therefore suffered from some dithering and a lack of an ability to take advantage of opportunities.

The final implementation did not suffer from these faults, but did lack the ability to combine preferences, and thus to choose compromise candidates. This is because of the system-level winner-take-all process which is central to the drives model.

### 10.6.2. Lorenz

The Lorenz scheme for action selection performed fairly badly, although the time-dependent internal variables did produce a notable 'goal-directedness' (that is, a tendency to concentrate on one system, or type of behaviour, for a length of time). Changes to the basic model were required to create ceilings on some reservoir levels, to change the relationship between consummatory actions and emptying of reservoirs, and to stop increases in reservoir levels at night when the animal is sleeping. But even with these changes the Lorenz ASM still performed fairly badly, in part because it proved impossible to calibrate both the increases in reservoir levels over time, as well as the decreases due to consummatory actions, so that the reservoir levels stayed in tune with the actual offsets from the homeostatic set-points. The reservoir levels for systems such as 'get water' would, with the right parameter settings, rise and fall appropriately at the beginning. However, the parameters could not be set so as to maintain a correspondence over different parts of the animal's lifespan (it will find water more easily as it learns the location of good water sources), and over a wide range of different instances of the SE (with different random distributions of water sources for instance).

With respect to the list of requirements in section 10.5, Lorenz's model has nothing to promote persistence of consummatory actions, does not deal well with all types



of sub-problem, and has a winner-take-all operation at the system level that means it cannot combine multiple preferences or choose compromise candidates. There is also an assumption that stimuli are summed.

### 10.6.3. Maes

The ASM that fared worst at this particular test of action selection was Maes' spreading activation network. This may be partly due to the fact that it was specified more precisely than any other mechanisms and so there was less room for favourable interpretations. But most of the shortcomings were due to several design faults, the main one being the decision to propagate motivational inputs to appetitive alternatives by way of consummatory alternatives. The problems caused by this decision become apparent when it is shown to be impossible to choose division rules that produce balanced competition between all sorts of nodes in all circumstances. With respect to the points in section 10.5, Maes' mechanism is good at producing interrupts when necessary and taking advantage of opportunities. However, it is not good at persistence (because the activation of the winning node is reset to zero and because there is nothing in the mechanism to take account of switching costs), and it is not good at preferring consummatory actions in one system over appetitive ones in another (partly due to the exponentially-increasing positive feedback between unexecutable consummatory nodes and executable appetitive nodes). While there is no system-level winner-take-all, there is also no combination of preferences and thus no means of selecting compromise candidates. Also, Maes' mechanism has binary-valued sensors of the environment and presupposes summation of stimuli.

Maes' ASM does not (as claimed in [Maes, 91a]) work well for animal-like action selection problems, with many competing goals/systems and many executable candidate nodes which have to be chosen amongst at each timestep.

#### 10.6.4. Rosenblatt & Payton

The most basic and naïve implementation of Rosenblatt & Payton's ideas, with summation of inputs and with no means of preferring consummatory over appetitive alternatives, fared rather badly at the test posed by the SE, but the final implementation, with a more sophisticated rule for combination of stimuli and with temporal and uncertainty penalties did best of all out of the four ASMs tested. It satisfies all of the requirements of section 10.5, with slight question marks only over the lack of tendency in the mechanism to promote persistence and contiguous action sequences. These points and several others are discussed in the next chapter, in which the extended Rosenblatt & Payton model is analysed extensively and is proposed as the best sort of method of calculating action selection.

### 10.7 Summary of Testing

Some of the important differences between how different mechanisms tackle the problem of action selection were shown in table 8.1. Table 10.1 below is similar to table 8.1, but instead compares each ASM that was tested against the list of requirements of section 10.5. Some of the  $\checkmark$ 's and  $\times$ 's in the table are somewhat debatable, but the table as a whole gives an indication of the differences between the mechanisms and the inadequacies that were discovered in them. The hierarchical decision structure (HDS) ASM in the results has not been introduced so far. It is an embodiment of the rigid switching type of hierarchy favoured by Tinbergen and Baerends, and is described in section 11.1.2, where it is contrasted with the type of hierarchy favoured by Rosenblatt & Payton.

The performance results for all mechanisms in all versions of the SE are summarised in table 10.2 and shown graphically in figure 10.3. The naïve implementation of the Rosenblatt & Payton ASM does not feature in either the table or the figure because its performance (genetic fitness) was always zero. This was because there was no tendency to prefer consummatory over appetitive options. It should be kept in mind

	Drives	Lorenz	HDS	Maes	NaïVe R&P	Extended R&P
All types of sub-problem?	×	×	×	×	✓	✓
Persistence?	?	✓	?	×	?	?
Activations $\propto$ current offsets?	✓	×	✓	✓	✓	✓
Consummatory before appetitive in same system?	✓	✓	✓	✓	×	✓
Consummatory before appetitive in other systems?	✓	?	✓	×	×	✓
Balanced competition?	?	?	✓	×	✓	✓
Contiguous action sequences?	?	✓	?	✓	?	?
Interrupts if necessary?	✓	✓	✓	✓	✓	✓
Opportunism?	✓	×	✓	?	✓	✓
No system-level winner-take-all?	×	×	×	✓	✓	✓
Combination of preferences?	×	×	×	×	✓	✓
Compromise candidates?	×	×	×	×	✓	✓
Real-valued sensors?	✓	✓	✓	×	✓	✓
Flexible combination of stimuli?	?	×	✓	×	✓	✓

Table 10.1: A table showing which of the six mechanisms tested in the SE satisfy the various criteria of section 10.5. '?'s signify that the answer is not very clear from the specification of the mechanism or the testing results.

		Standard SE	SE v1.	SE v2.	SE v3.	Average
<b>Drives</b>	av. performance ( $\bar{x}$ )	6.44	3.29	6.41	8.78	<b>6.23</b>
	std. dev. ( $s$ )	6.42	3.89	6.55	8.17	6.26
	num. tests ( $n$ )	1645	1721	1627	1607	total no. = 6600
	std. dev. of $\bar{x}$ ( $s_{\bar{x}}$ )	0.16	0.09	0.16	0.20	0.15
<b>Lorenz</b>	av. performance ( $\bar{x}$ )	2.39	1.40	2.48	4.58	<b>2.71</b>
	std. dev. ( $s$ )	2.34	1.51	2.38	4.82	2.76
	num. tests ( $n$ )	1654	1653	1625	1668	total no. = 6600
	std. dev. of $\bar{x}$ ( $s_{\bar{x}}$ )	0.06	0.04	0.06	0.12	0.07
<b>HDS</b>	av. performance ( $\bar{x}$ )	6.69	3.48	6.48	7.79	<b>6.11</b>
	std. dev. ( $s$ )	6.24	3.77	6.39	7.43	5.96
	num. tests ( $n$ )	1641	1679	1693	1587	total no. = 6600
	std. dev. of $\bar{x}$ ( $s_{\bar{x}}$ )	0.15	0.09	0.16	0.19	0.15
<b>Maes</b>	av. performance ( $\bar{x}$ )	0.16	0.22	0.18	0.43	<b>0.25</b>
	std. dev. ( $s$ )	0.41	0.49	0.44	0.66	0.50
	num. tests ( $n$ )	1692	1625	1672	1611	total no. = 6600
	std. dev. of $\bar{x}$ ( $s_{\bar{x}}$ )	0.01	0.01	0.01	0.02	0.01
<b>Extended R&amp;P</b>	av. performance ( $\bar{x}$ )	8.09	3.61	8.16	13.38	<b>8.31</b>
	std. dev. ( $s$ )	7.06	3.63	6.66	9.16	6.63
	num. tests ( $n$ )	1649	1693	1653	1605	total no. = 6600
	std. dev. of $\bar{x}$ ( $s_{\bar{x}}$ )	0.17	0.09	0.16	0.23	0.16

Table 10.2: The average genetic fitnesses for the final implementations of the five ASMs tested in the four versions of the SE described in section 9.1.1. For each ASM, each test in a version of the SE consists of  $n$  runs, which have a mean performance of  $\bar{x}$  and a standard deviation of  $s$ . The expected standard deviation of  $\bar{x}$  (i.e.  $s_{\bar{x}}$ ) is equal to  $(s/\sqrt{n})$ .

that there are several rather arbitrary components to the testing process which gives rise to the exact performance measures. Therefore the existence of a small difference between the performance of two ASMs is a rather indefinite indicator that one is better than the other. This is in contrast to the hard and fast results concerning gross deficits in performance (e.g. section 9.3 and table 10.1).

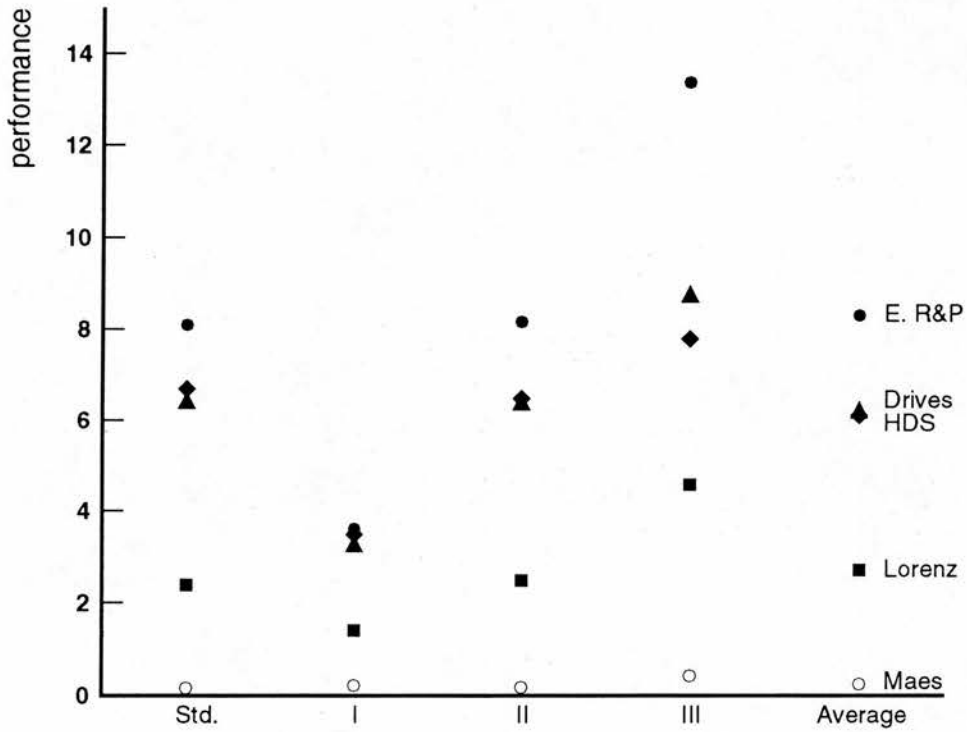


Figure 10.3: A graph showing the performance measurements (genetic fitnesses) of the final implementations of the ASMs that were tested in the SE. **Std** stands for the normal SE, **I** stands for the first altered version of the SE, **II** stands for the second altered version of the SE and **III** stands for third altered version of the SE.

## Chapter 11

# How Best to Compute Action Selection

In chapter 9 several ASMs were tested and the extended Rosenblatt & Payton ASM was seen to perform better than the others. This ASM is a development of Rosenblatt & Payton's basic ideas, to which temporal and uncertainty penalties were added by the author, and for which a more specific rule for combination of preferences was produced. This ASM (with the extensions) has no serious deficits and satisfies all of the criteria of section 10.5 (albeit with a couple of question marks over persistence and contiguous action sequences). In this chapter the extended Rosenblatt & Payton ASM is proposed as a sensible approach to action selection, and is analysed in more detail. The proposals are rather preliminary; only a broad outline of the suggested optimal sort of mechanism is given here, together with some suggestions for future investigation. The suggestions are intended as a well-founded platform for further development, rather than the last word in how action selection should be computed. However, each of the different design decisions that have been made for this extended Rosenblatt & Payton mechanism are considered and argued for at length.

The structure of this chapter is as follows: *(i)* the theory and design decisions behind the suggested optimal ASM are examined in detail, *(ii)* some examples of the operation

of this type of ASM are given, and (iii) some suggestions are given as to how these preliminary ideas may be extended and improved upon. Appendix A also contains diagrams showing the implementation of each system of this ASM for the problem posed by the SE.

## **11.1 Theory Behind the New Mechanism**

This section now looks at some of the more important theoretical design decisions that have been made for the extended Rosenblatt & Payton mechanism. Some of these points have been covered to some extent already, and are now covered again but in greater depth.

### **11.1.1 Hierarchy or No Hierarchy?**

The term hierarchy was defined in section 7.3. There are three reasons to prefer a hierarchical over a non-hierarchical mechanism:

1. The action selection problem for real animals may well be intrinsically hierarchical (as the diagrams of Baerends suggest in figures 8.8 and 8.9), and it would seem likely that the best solutions to the problem should reflect this.
2. Non-hierarchical mechanisms do not allow combination of preferences from higher-level layers to lower-level ones.
3. All of the non-hierarchical mechanisms tested here (drives, Lorenz, and Maes) performed less well in the SE than the free flow hierarchy of the extended Rosenblatt & Payton mechanism.

### 11.1.2 Hierarchical Decision Structure or Free-Flow Hierarchy?

If it is accepted that a hierarchical mechanism is preferable, then what sort of hierarchy should it be? This section distinguishes between two types of hierarchy — the *hierarchical decision structure* and the *free-flow hierarchy* — and goes on to explain why the latter is more desirable, even though most traditional hierarchies belong to the former category.

#### Hierarchical Decision Structures

The defining characteristic of hierarchical decision structures is that there is a winner-take-all operation at *every* level, from the system level down. A *decision* is made at each level (by whatever process) as to which system or sub-system or action is most appropriate, and only that particular node is allowed to remain active and to pass activation down the hierarchy. Another name for this sort of mechanism is a *rigid switching mechanism*.

Two examples of this sort of hierarchy are from Tinbergen and Baerends (see sections 8.2 and 8.4). In Tinbergen's model there is some unspecified form of intra-layer inhibition so that "centres of the same level mutually suppress each other's activities". In Baerends' models (figures 8.8 and 8.9) system-level nodes inhibit each other so that only one survives and nodes at lower levels control the activity they send out so that only one of their subordinates receives activation.

#### Free-Flow Hierarchies

A free-flow hierarchy (such as Rosenblatt & Payton's ASM) does not implement a switching process at every level, but rather places no restrictions on the flow of information or activation through the hierarchy. An exception to this is at the level of the behavioural final common path, where a winner-take-all process must occur to decide which of the actions will be executed.



An analogy can be made between the two hierarchical schemas for action selection considered here and the decision-making of a group of ministers or government officials. Imagine that each official is responsible for a particular aspect of policy, such as education, foreign affairs, unemployment or the economy. Imagine further that the group are meeting to decide upon the future political strategy of their nation, and that each official is constrained to evaluate the various strategy options from the standpoint of their individual department only. The prime minister or president presides over the council and makes the ultimate decision as to which of the alternative options gets chosen.

Decision-making with a hierarchical decision structure is equivalent to the prime minister or president listening to each of the various officials in turn and then selecting the one department for which the decision is most crucial. The future policy is then chosen to be that which is most beneficial to the particular department.

By contrast, decision-making with a free-flow hierarchy is analogous to the prime minister or president listening to each official outline the pros and cons of each action for their particular aspect of government, and then coming to a decision which takes into account the views of all officials simultaneously. So, for instance, the official in charge of unemployment might express a strong disapproval of option A because of the number of people who would be put out of work, but be ambivalent as to which of the other options is chosen. The official for foreign affairs could mildly favour option B because it would ease diplomatic relations with another country, but express no preference or dislike for other options. And the various options might each be favoured to differing degrees by the official concerned with the economy, in accordance with the likely effects on the national finances. The preferences of each official for each policy option can be combined to calculate which option is most beneficial when all aspects are taken into consideration.

## Reasons for Preferring Free-Flow Hierarchies

Free-flow hierarchies are preferable for action selection because they enable combination of preferences and compromise candidates. For instance, in figure 10.1, with a free-flow hierarchy the animal can choose the action 'move west' to approach the food *and* move away from the hazard, whereas with a hierarchical decision structure the animal will choose the action 'move south'. Also, there is no loss of information in a free-flow hierarchy since there are no restrictions as to how many nodes can remain active.

In addition to these obvious differences, Maes [91a] and others have criticised traditional hierarchical-type mechanisms for their lack of robustness and the top-down, preprogrammed sort of control that they exert. In a hierarchical decision structure the initial decision as to which system to choose is vitally important. If this decision is made incorrectly then the whole performance can collapse. This leads to a lack of robustness. A free-flow hierarchy does not make any decisions until the lowest level and so there is a much more distributed type of processing involved, with a consequent higher degree of robustness.

Hierarchical decision structures have also been criticised because of the '*sensory bottleneck*' that occurs. In order to make correct initial decisions in a hierarchical decision structure, the top level has to receive and process all, or most of the sensory information relevant to the lower layers, leading to the sensory bottleneck (see figure 11.1).

For any system of an action selection mechanism there are several relevant internal, indeterminate and external stimuli. Some of these stimuli may be relevant to a whole system or sub-system (e.g. 'perception of partially built nest' to the nest building system), whereas others will be relevant to only one specific action (e.g. 'perception of egg outside nest' to egg retrieval). The desirability of, or tendency for, carrying out any specific action will usually depend on several stimuli, some of which (e.g. external stimuli) are relevant only to that action, and some of which (e.g. internal, motivational

stimuli) have more widespread relevance, and are relevant to many other actions as well [e.g. Hinde, 1953]. In a hierarchical decision structure the desirability of each system needs to be calculated, so that the systems can be properly selected between. This desirability ('drive strength') for each system will be equal to the maximum of the desirabilities of each of the actions of the system, and should be calculated using 'releasing stimuli' relevant to the actions in that system (such as the presence of water for the 'get water' system) as well as motivational stimuli (e.g. water deficit). This issue was discussed in section 9.3.1, where it was noted that in order to produce opportunism, persistence and contiguous action sequences, the drive strengths needed to be calculated as a function of both cue and motivational stimuli. As an example, the 'get food' drive strength needed to be a function of both the internal food deficit and the external food stimulus (which will be strongest if there is food in the animal's square, weaker if there is food that is perceived or remembered, and zero otherwise).

The desirability of choosing a particular system, sub-system, or action is a function of both the need for that sort of behaviour, and the appropriateness of that sort of behaviour in the current situation (for instance, the 'get water' system is less useful if there is no water in the animal's square, no water perceived, and only a very weak memory of a water source). Therefore, the system-level nodes must each receive all, or at least most, of the relevant stimuli for all of the actions in their system, which produces the sensory bottleneck mentioned above. Since there is no selection between higher-level nodes in a free-flow hierarchy, it does not matter that the system or sub-system node activations are not exactly proportional to the desirabilities of selecting that system or sub-system, as long as the activations of the *action* nodes are correctly calculated.

It is apparent from the arguments above that there are several *computational* reasons for preferring free-flow hierarchies. Two other types of supporting argument are now given.

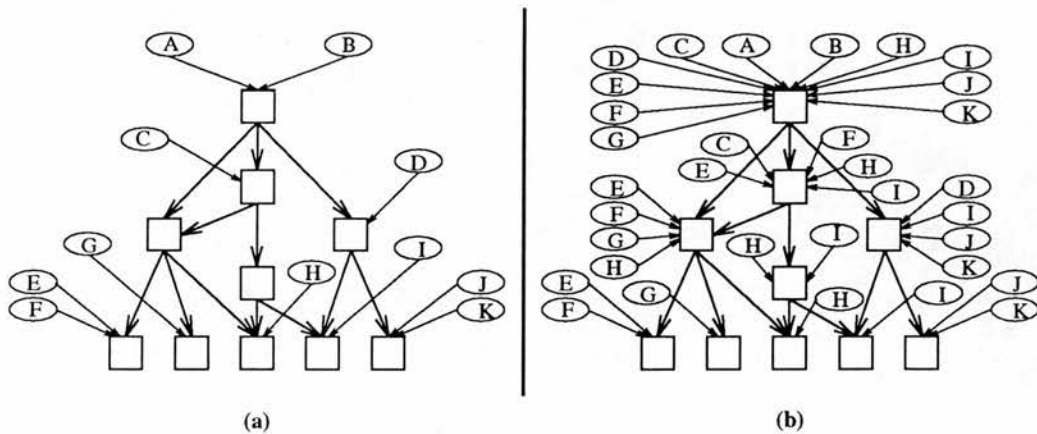


Figure 11.1: 'Sensory bottleneck' — different stimuli are relevant to different subsets of the whole system. In a free-flow hierarchy, these stimuli can impinge on the relevant nodes (as shown in (a)). In a hierarchical decision structure, higher-level nodes need information about the stimuli to nodes that they are superior to, in order for the higher-level nodes to be selected between appropriately (as shown in (b)).

### Animal Behaviour Observations Supporting Free-Flow Hierarchies

Evidence from animal behaviour is now presented in support of the claim that the degree of preference for respective alternatives is maintained throughout each system, rather than the winning system shutting down all the others. These three examples come from pages 243–245 of [Lorenz, 81].

1 : The simultaneous firing of antagonistic muscles due to simultaneous excitation of conflicting systems:

Additive superposition is found even in cases in which the two independent motivations activate antagonistic muscles. ... A conflict between motivations in *Anatidae*, one demanding a forward extension of the neck, the other a retraction, which can occur in a goose wanting to eat grain offered in a

human hand and not quite daring to do so, produces a violent trembling of the neck. . . . The “trembling neck” of the goose probably is effected by conflicting innervation of the antagonistic muscles, but we cannot be certain about this.

2 : The simultaneous operation of opposing sets of fins due to the strong activation of conflicting systems:

In disputed territory, certain cichlids (*Etroplus maculatus*) position themselves opposite another, threatening across the border separating their territories. As in every threat, aggressive motivation is contending with that for escape. Whenever one of the adversaries moves a short distance forward into enemy territory, it appears as if he were swimming into a current, the speed of which rapidly increases as one proceeds upstream. This effect is produced by the action of the pectoral fins, which are sculling in reverse, and doing so more and more intensely the farther the fish moves into the other’s territory. The tail fin is under the control of aggressivity and the pectorals under that of escape, and the observer cannot help feeling – ridiculously – that the pectorals are more afraid than the tail, because they are nearer to the enemy.

3 : The facial expressions of a dog can simultaneously display the intensity of two separate, conflicting motivations, fear and aggression (figure 11.2).

These three examples all serve to illustrate the point that the activation levels of different nodes in a hierarchy are likely to be maintained, as in a free-flow hierarchy, rather than only one at each level surviving, as in the winner-take-all operation of a Tinbergen-like hierarchical decision structure. As well as the three examples stated above, Lorenz went further to state that

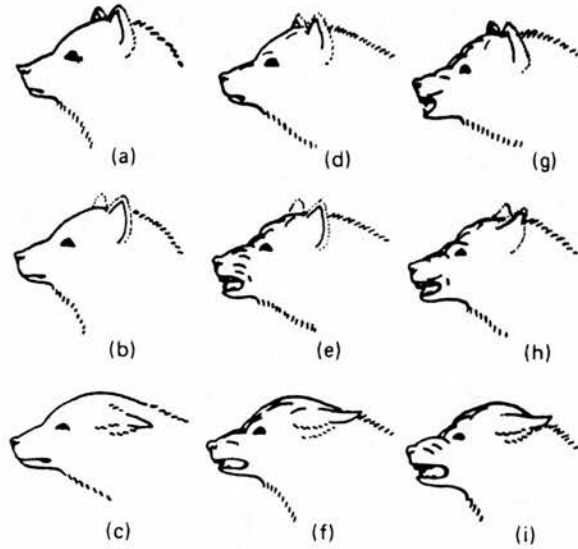


Figure 11.2: Facial expressions of a dog showing differing degrees of fear and aggression. Readiness to flee increases from top to bottom, aggression increases from left to right. Taken from [Lorenz, 81].

In nature, behaviour activated by a single motivation is found at least as rarely as hybrids differing in only one gene. A higher animal in its natural habitat must always be ready to undertake a great number of different and – as often as not – mutually exclusive actions, and what it finally does is almost always a compromise made among several necessities.

### SE Test Results Supporting Free-Flow Hierarchies

In order to make a stronger test of the hypothesis that free flow hierarchies are better for action selection than hierarchical decision structures, a hierarchical decision structure was implemented and tested in the SE so that the performance results could be compared with those for the extended Rosenblatt & Payton ASM (section 9.3.4). The hierarchical decision structure (HDS) which was used was not a direct implementation of either Tinbergen's or Baerends' ideas, both of which are somewhat imprecise and unclear, but instead was an implementation of the generic concept.

The HDS that was tested was *not* obtained using the same motivational inputs as the extended Rosenblatt & Payton ASM, because this had been optimised to take



account of combinations of preferences. Instead, the system-level node activations were calculated as in the drives model, which had already been proven to work well for selection between systems. The connections to lower-level nodes (all nodes except system-level nodes) were the same as in the extended Rosenblatt & Payton ASM. The same nodes, connections and weights were retained, including the temporal and uncertainty penalties. In order to make sure that preferences were not combined from sub-system nodes in the same system onto action nodes, the combination rule for each node was changed from

$$A_j = \left( \frac{\max_i(P_{ij}^+) + \alpha \sum_{i=1}^{N^+}(P_{ij}^+)}{1 + \alpha} \right) + \left( \frac{\min_i(P_{ij}^-) + \beta \sum_{i=1}^{N^-}(P_{ij}^-)}{1 + \beta} \right)$$

to

$$A_j = \max_i(P_{ij}^+)$$

where  $P_{ij}^+$  and  $P_{ij}^-$  are the positive and negative preferences from node  $i$  for node  $j$ ,  $N^+$  and  $N^-$  are the numbers of such preferences for node  $j$ , and  $\alpha$  and  $\beta$  are constants. That is to say, the activation of each node was calculated as the maximum of the positive inputs to each node (or the maximum of multiplied pairs where there were associated pairs of external and motivational stimuli). Neither combination of preferences, nor negative preferences, were allowed. So, for instance, the activation of the 'move fast north' node was calculated as

$$A_j = \max( ('run\ away\ from\ p1\ square' \times 'opp\ p1\ [north]') + ('leave\ current\ square\ because\ of\ p1' \times 'all\ directions\ [north]') + ('run\ away\ from\ p2\ square' \times 'opp\ p2\ [north]') + ('leave\ current\ square\ because\ of\ p2' \times 'all\ directions\ [north]') )$$

The results of testing the HDS in the four different versions of the SE then the average results ( $\bar{x}$ ) were 6.69, 3.48, 6.48 and 7.79, with an average over the four of **6.11**. The standard deviations ( $s$ ) of the individual measurements, in the tests in each of the four versions of the SE, were 6.24, 3.77, 6.39 and 7.43; and for the average result over all four versions was 5.96. The expected standard deviations of the sample averages ( $s_{\bar{x}} = (s/\sqrt{n})$ ) were 0.15, 0.09, 0.16 and 0.19; and for the average result over all four versions was 0.15.

$$\left( \frac{|(8.31 - 6.11)|}{\sqrt{\frac{69.6}{6600} + \frac{37.3}{6600}}} \right) = 13.4$$

which is greater than 3.291.

The issue of what sort of hierarchy to choose has been discussed at length, and several reasons for preferring free-flow hierarchies have been explained. Computational arguments, ethological arguments and the results of tests in the SE all support this same conclusion.

### 11.1.3 How to Combine Preferences

It was mentioned in sections 9.3.4 that simple summation of preferences for a node, i.e. where

$$A_j = \sum_i (P_{ij}^+) + \sum_i (P_{ij}^-)$$

is not optimal in most cases. The reasons for this are now explained and consideration is given as to how preferences should be combined.

The concept of combination of preferences relies on the idea that the relative values of node activations bear some relationship to the future expected genetic fitness, or 'utility', of those nodes. So, for instance, the activation level of the 'avoid predator1s'



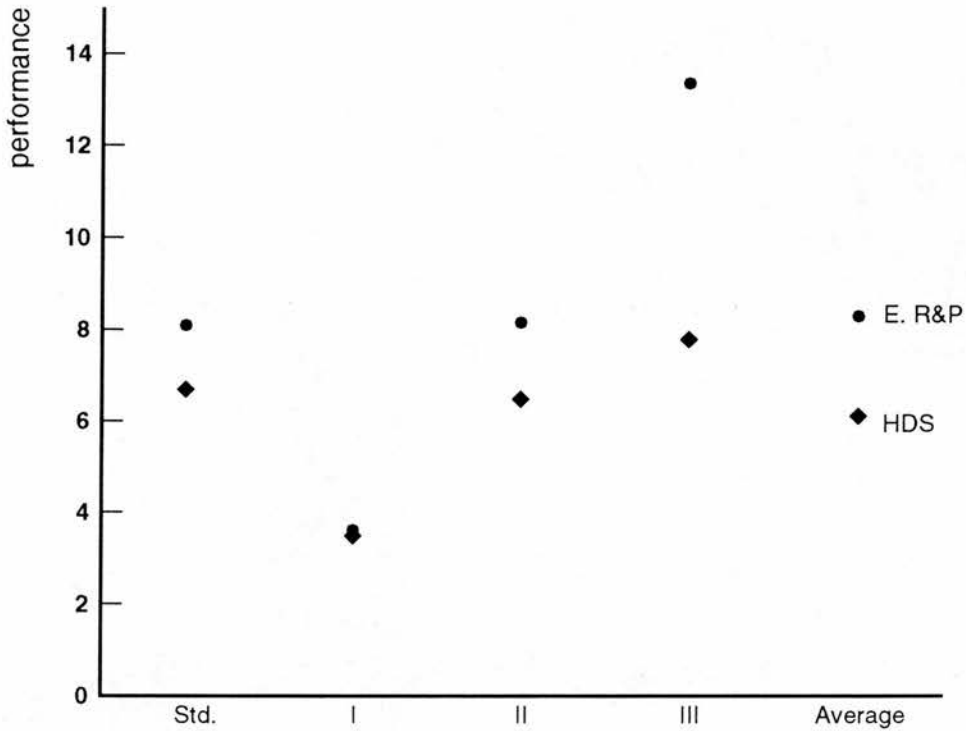


Figure 11.3: A graph comparing the performance measurements (genetic fitnesses) of a hierarchical decision structure (HDS) ASM and a free flow hierarchy ASM (i.e. the extended Rosenblatt & Payton (E. R&P) ASM). **Std** stands for the normal SE, **I** stands for the first altered version of the SE, **II** stands for the second altered version of the SE and **III** stands for third altered version of the SE.

system-level node is proportional to the increase, or lack of decrease, in future expected genetic fitness that can be obtained by trying to avoid predators (by whatever means). As another example, the activation of the 'look around' action node is proportional to the increase, or lack of decrease, in future expected genetic fitness that can be obtained by performing the action LOOK\_AROUND.

Assuming that system-level nodes have appropriate activation levels then lower-level nodes will also have appropriate activation levels as long as activation is passed down

to them correctly and as long as the different inputs to the nodes are combined correctly. But if preferences are combined incorrectly then the lower-level activations will no longer be in proportion to the utility of the nodes and the whole mechanism will then start to produce sub-optimal action selections.

There are two sorts of situation in which strict summation is the wrong way to combine preferences, and these explain why the performance of the first, naïve implementation of Rosenblatt & Payton ASM was so poor (section 9.3.4).

### **1. Recombining Preferences Originating from the Same System**

Imagine the case in figure 11.4. Here the action node 'move west' is receiving preferences from 'approach perceived food', 'approach remembered food' and indirectly from 'explore for food', but these three inputs are all originally triggered by the same motivational stimulus. The utility of proceeding in a particular direction in order to approach perceived food, approach remembered food and explore simultaneously is not equal to the sum of the utilities of each individually. If the animal encounters the perceived food first then it will very possibly become satiated and should not continue to approach the remembered food or explore afterwards.

The problem is somewhat similar to that experienced in Maes' network (figure 9.9), in which satisfactory division rules for inputs to appetitive nodes cannot be devised because the inputs can arise from several goals or from one. In a free-flow hierarchy there is no problem with feeding motivational information from the system-level node to consummatory and appetitive nodes at the next lowest level, but there is a problem when passing preferences further down the hierarchy, as shown in figure 11.5. The dilemma this time is that it is not possible to know whether the preferences being combined come from the same system or different systems. The problems in this case are somewhat less severe, because fairly competent action selection can still take place with limited combination of preferences or with none at all, whereas the proper transfer of goal- or motivation-type information is critical.

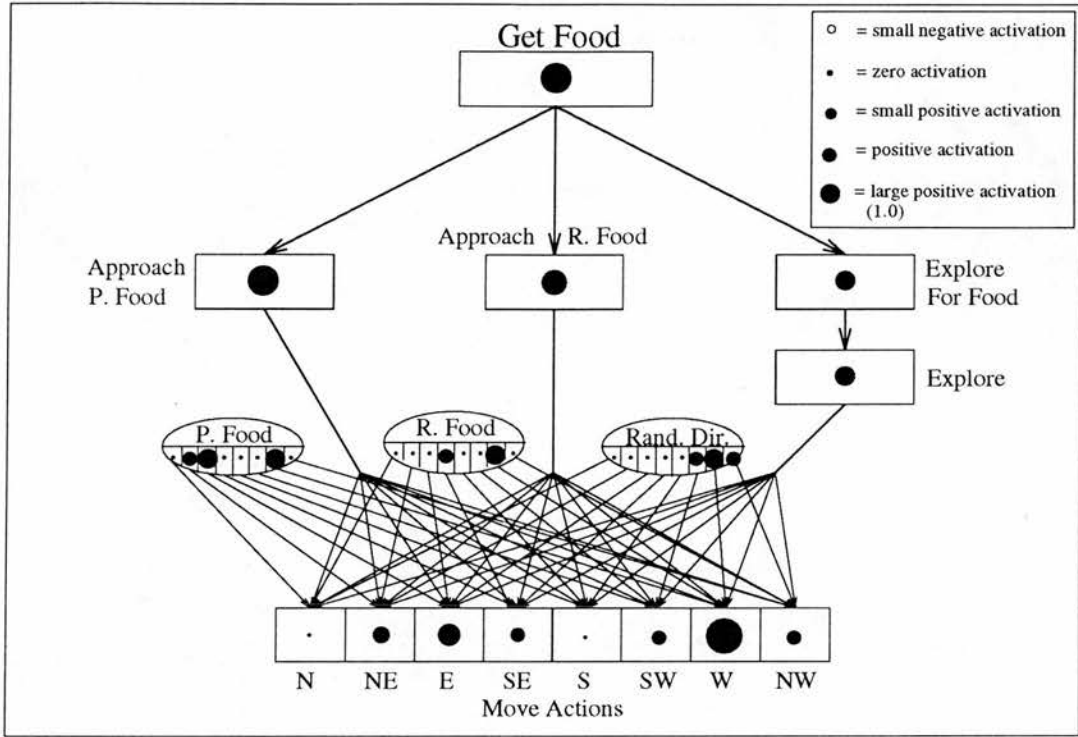


Figure 11.4: Preferences from the same system onto the same action node. MOVE.WEST receives excitation from 'approach perceived food', 'approach remembered food' and 'explore for food'. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

## 2. Combining Preferences from Appetitive Nodes

Consider the situation in figure 11.6. Here the action 'move north-east' receives preferences from the nodes 'approach receptive mate', 'approach perceived water' and 'approach shade'. The three preferences all originate from different systems in this case, but the utility is still not equal to the *sum* of the preferences from the three nodes. An appetitive alternative has utility because it brings about a situation in which a consummatory action can be performed. But because of the constant possibility of interruption of any appetitive/consummatory sequence

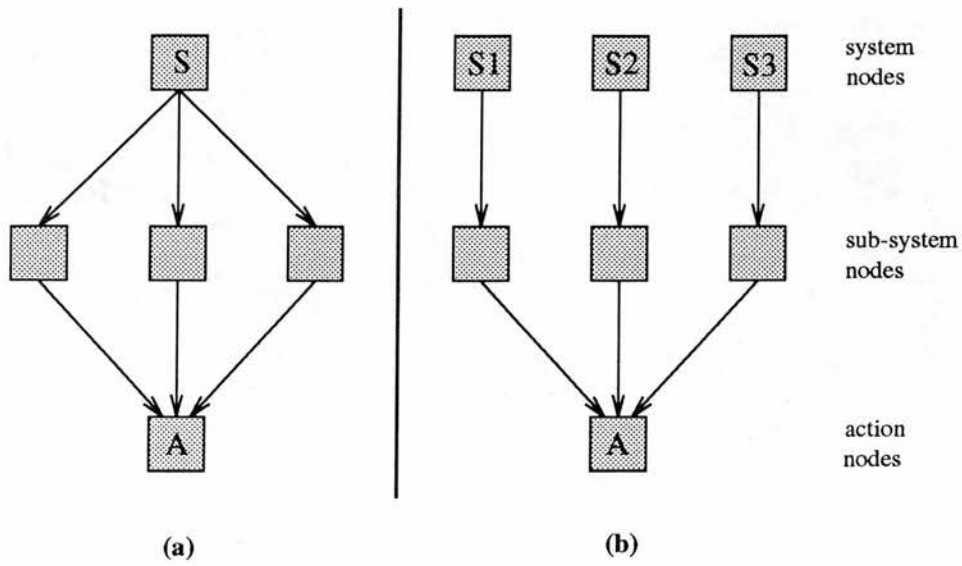


Figure 11.5: The dilemma when combining preferences: in (a) the preferences come from the same system, in (b) they come from different systems.

(by the need to avoid predators or some other urgent demand) the utility of performing three appetitive actions simultaneously is much less than the sum of the utilities of performing them each individually. This is because the probability of performing the joint appetitive action and then eventually continuing to perform consummatory actions for all three systems is fairly remote. For instance, by the time the animal has approached, courted and copulated with the mate, and has then approached and drunk the water (assuming it gets to do so) then the probability of its still being uninterrupted and able to approach the shade is fairly low. In any case, its body temperature may have improved of its own accord by then, making it no longer necessary to approach shade. In general, the longer the likely gap between an appetitive and consummatory action, the less likely that the one will lead to the other (because of the dynamic, rather unstable and unpredictable nature of the environment), and so the lower the expected value of performing the appetitive action. For this reason, an action which receives preferences from

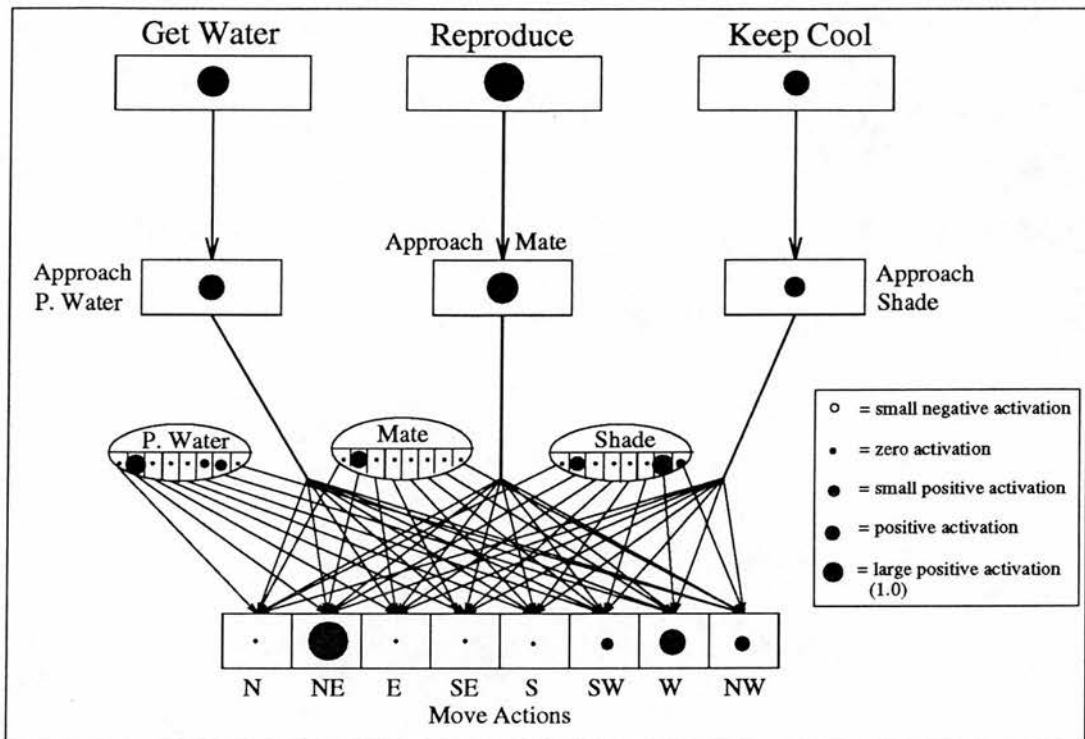


Figure 11.6: Preferences from appetitive nodes in different systems to the same action node. MOVE\_NORTH\_EAST receives preferences from 'approach perceived water', 'approach perceived mate' and 'approach perceived shade'. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

several appetitive nodes should have a utility which is significantly less than the sum of all the individual preferences for it.

For these two reasons a simple summation of preferences is not adequate. For the extended Rosenblatt & Payton ASM a tradeoff was implemented which drew a balance between a simple sum of the inputs and the use of the maximum-valued input, i.e:

$$A_j = \left( \frac{\max_i(P_{ij}^+) + \alpha \sum_{i=1}^{N^+}(P_{ij}^+)}{1 + \alpha} \right) + \left( \frac{\min_i(P_{ij}^-) + \beta \sum_{i=1}^{N^-}(P_{ij}^-)}{1 + \beta} \right)$$

where  $P_{ij}^+$  and  $P_{ij}^-$  are the positive and negative preferences from node  $i$  for node  $j$ ,  $N^+$  and  $N^-$  are the numbers of such preferences for node  $j$ , and  $\alpha$  and  $\beta$  are constants. With values of  $\alpha = \frac{1}{(N^+)^2}$  and  $\beta = \frac{1}{(N^-)^2}$  this produced reasonable results, as described in section 9.3.4. It may prove to be possible to develop a more principled scheme for dealing with the two difficulties just enumerated, but this was not attempted in this work.

#### 11.1.4 Selecting Between Consummatory and Appetitive Alternatives

In the basic and naïve Rosenblatt & Payton hierarchy there was no differential weighting towards consummatory over appetitive alternatives. Thus, if both 'approach remembered water' and 'eat food' are relevant to the current situation (the animal can remember water and there is food in its square), then the animal is equally likely to choose either alternative (assuming equal-valued stimuli) despite the fact that the latter requires lower expenditure of time to achieve that reward, and also gives a more immediate, and therefore more definite, reward. When the Rosenblatt & Payton ASM, complete with the combination of preferences rule just described, but with no differential weightings, was tested in the SE then the resulting genetic fitness was invariably zero because the animal hardly ever chose consummatory actions such as MATE over appetitive actions such as the MOVE ones.

There needs to be a tendency for consummatory over appetitive options, and also for more immediately rewarding over less immediately rewarding appetitive options (e.g. for 'approach perceived food' over 'explore for food'). As mentioned in section 10.5, there are two main concerns that need to be taken into account when deciding how to build in the tendency towards actions that maximise the expected benefit per unit of time expended:

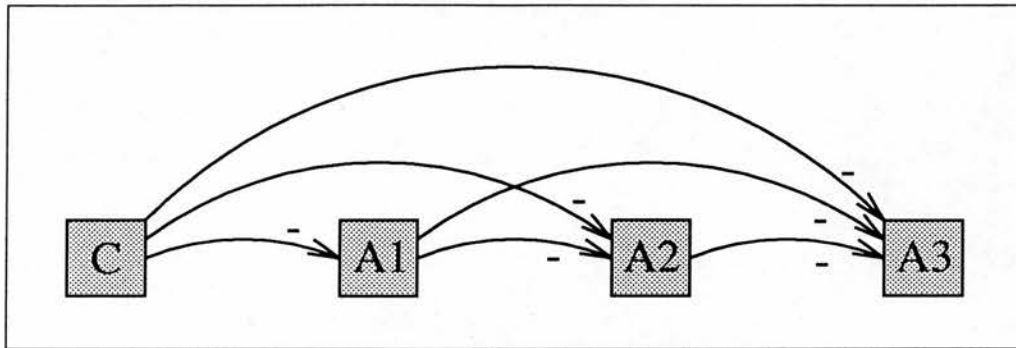


Figure 11.7: Arbitration between alternatives in the same system using intra-layer inhibitory connections to less immediate alternatives. If 'C' is active it inhibits all other nodes in its system, if 'A1' is active it inhibits 'A2' and 'A3', and so on.

1. if any two alternatives in the *same* system are both applicable to the current situation then the consummatory option, or the appetitive one 'closest' to it, must tend to dominate over less immediately and less certainly rewarding options.
2. if any two alternatives in two *different* systems are competing then, other things being equal, the one closest to the consummatory alternative in its own system should be chosen.

One solution to the proper production of sequences is that shown in figure 11.7, in which, if it is applicable to the current situation, any node suppresses all of those nodes that are further from the consummatory one. This scheme is not used here because, although it satisfies (1) above, it does not satisfy (2) (see figure 11.8).

The scheme used for the extended Rosenblatt & Payton ASM, as explained in section 9.3.4, is a development of Sutton & Barto's idea of *temporal discount factors* (see section 8.8.5). It is a scheme of differential inhibition of consummatory and appetitive nodes according to the extra time expenditure required for, and degree of uncertainty of, a

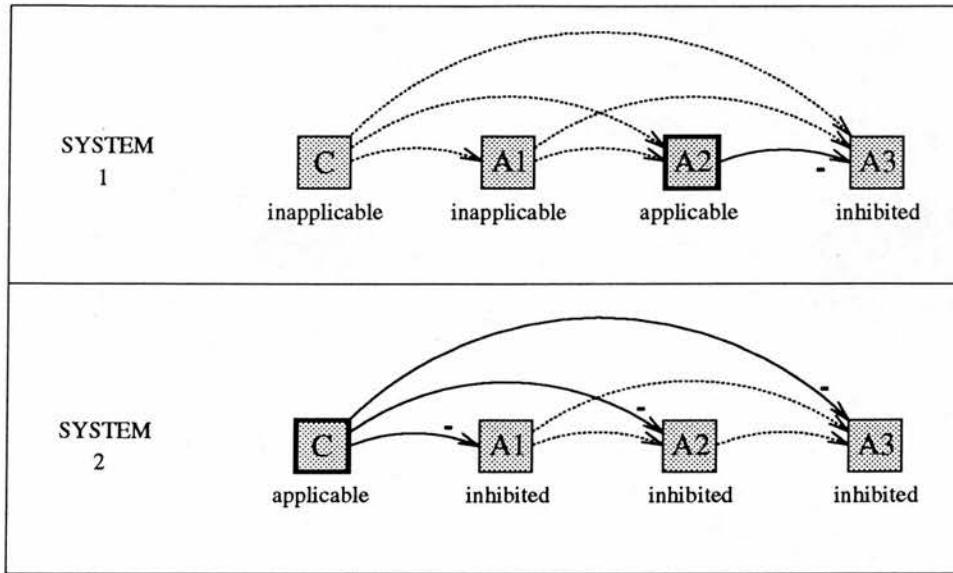


Figure 11.8: Intra-layer inhibitory connections do not produce balanced competition between alternatives in different systems. 'C' in system 2 is not preferred to 'A2' in system 1.

real benefit being obtained. Figure 11.9 shows the scheme, and that it does indeed provide a solution to both (1) and (2) above.

It should be noted that the scheme in figure 11.9 adheres to the same general policy of free flow of information as the basic Rosenblatt & Payton hierarchy. Appetitive options are penalised because they are less certain to lead to a benefit to the animal (the sequence may be interrupted by more urgent needs such as the need to avoid predators), and also because they require the animal to use more of its time in order to gain the benefit. Therefore appetitive options in general are less beneficial to the future expected genetic fitness than consummatory actions; they are less 'worthwhile'. By biasing against appetitive nodes, rather than shutting them down, there is still flexibility and the opportunity to express preferences from appetitive nodes, albeit with a weaker influence. On occasion a consummatory action will not be selected even



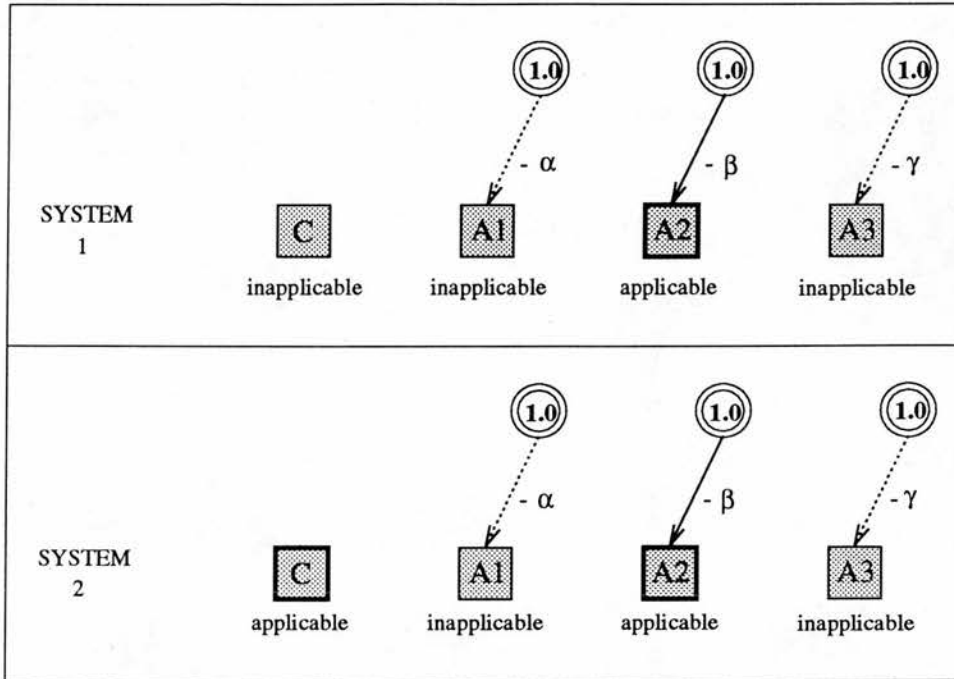


Figure 11.9: Inhibition from penalty nodes enables balanced competition between alternatives in the same system and alternatives in separate systems. All other things being equal, 'C' in system 2 would be preferred to both 'A2' in system 2 and 'A2' in system 1.

if it is applicable to the current situation. For instance, the animal may approach a large water source which is nearby to a large food source rather than drink from the small water source which it is at.

## 11.2 Examples

This section now takes a few of the systems in the ASM and demonstrates how activation might flow into them from the different types of stimuli (internal, external and indeterminate), and then how activation might be passed down through the nodes in the hierarchy to accumulate in the action-level nodes. The first three examples show

the passages of activation in three individual systems and the fourth example shows the passage of activation through three systems simultaneously and how they interact.

**'stay close to cover' –**

Figure 11.10 shows an example of how activation in this system can be propagated through the hierarchy in order to pass the right amounts of excitation to the appropriate action nodes. Note how simple summation of inputs would cause too much activation to be passed to the 'move east' and 'move south' nodes.

**'avoid predator1s' –**

Figure 11.11 shows activation passing through the 'avoid predator1s' system. This system is urgent, and so passes large activations to the action nodes when the relevant stimuli are present.

**'get food' –**

Figure 11.12 shows activation passing through the 'get food' system. The temporal and uncertainty penalties restrict the influence of the appetitive nodes, but allow them to beat consummatory nodes if necessary. In this example, the perception of a large-valued food source to the west has produced a greater activation in MOVE.WEST than the perception of a small-valued cereal food source in the animal's square has in EAT.CEREAL.FOOD.

**'reproduce', 'sleep in den' and 'keep clean' –**

To give some idea of the dynamics of the whole mechanism working together, three systems are shown simultaneously in figure 11.13. The system 'keep clean' is able to cause the action CLEAN to be selected because there are no strong stimuli for the 'reproduce' system and no great need for the 'sleep in den' system.

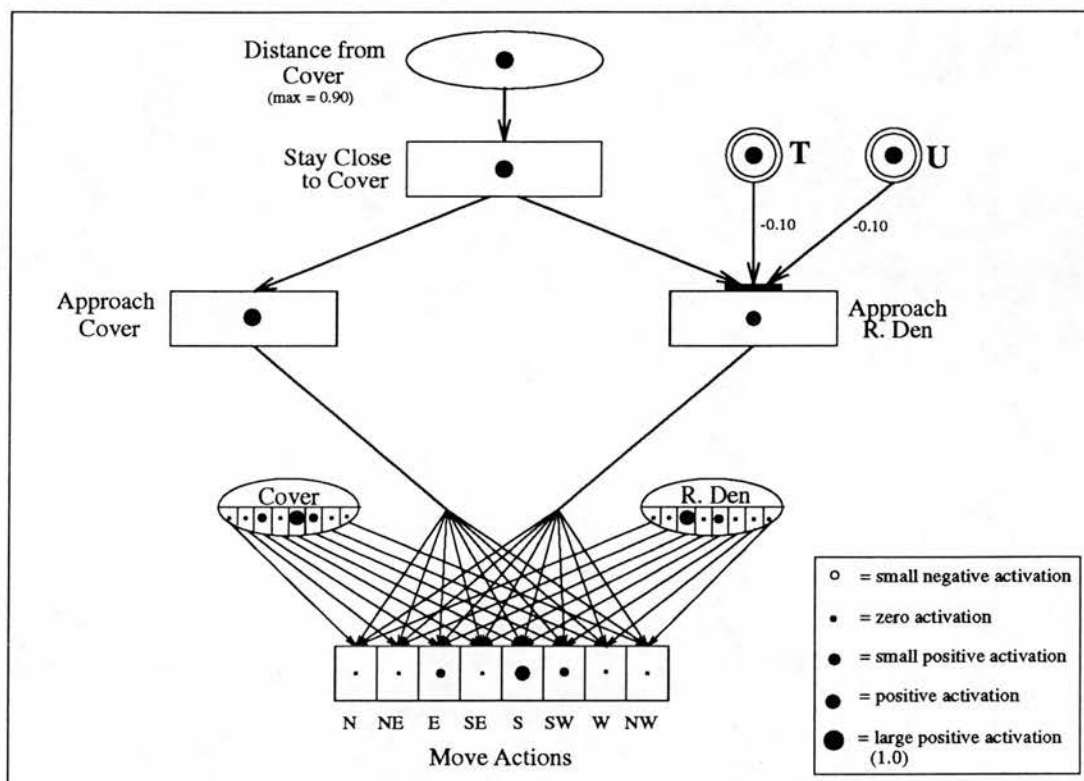


Figure 11.10: An example of the passage of activation through the 'stay close to cover' system. The possible effects of other systems on the action node activations are not considered. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

### 11.3 Discussion

The theory and the most important design decisions behind the extended Rosenblatt & Payton ASM have been explained at length here. The mechanism is fairly simple: a free-flow hierarchy with a more complicated rule for combination of preferences and with inhibitory connections to enable temporal and uncertainty penalties. This framework is fairly basic but is theoretically well-founded and is still able to outperform the other types of ASM tested in the SE.

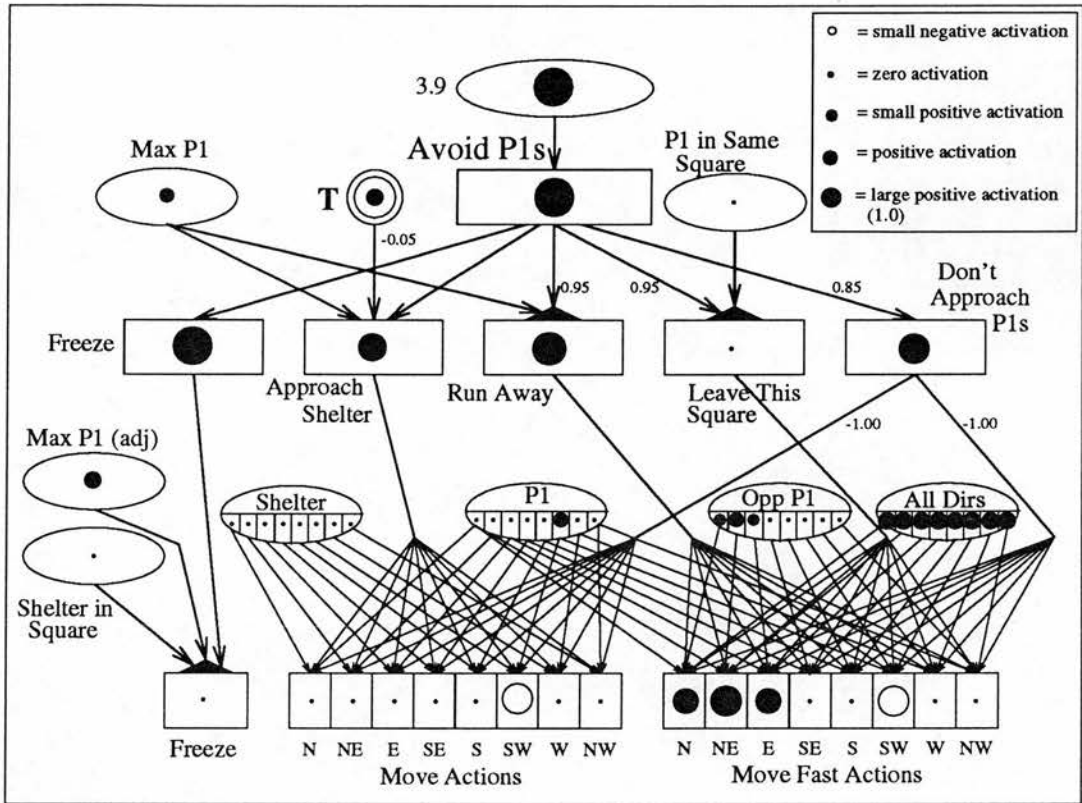


Figure 11.11: An example of the passage of activation through the 'avoid predator1s' system. The possible effects of other systems on the action node activations are not considered. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

This mechanism has been derived in part from ideas from both robotics and ethology. Its design though has been motivated more by the desire to optimise computational performance than the need to account for animal behaviour phenomena. There is no discussion here as to whether phenomena such as displacement activity can be observed in the behaviour produced by the extended Rosenblatt & Payton ASM in the SE. Similarly, the design of the novel ASM has not been influenced by the desire to produce a biologically realistic model, with brain-like neurons and synapses.

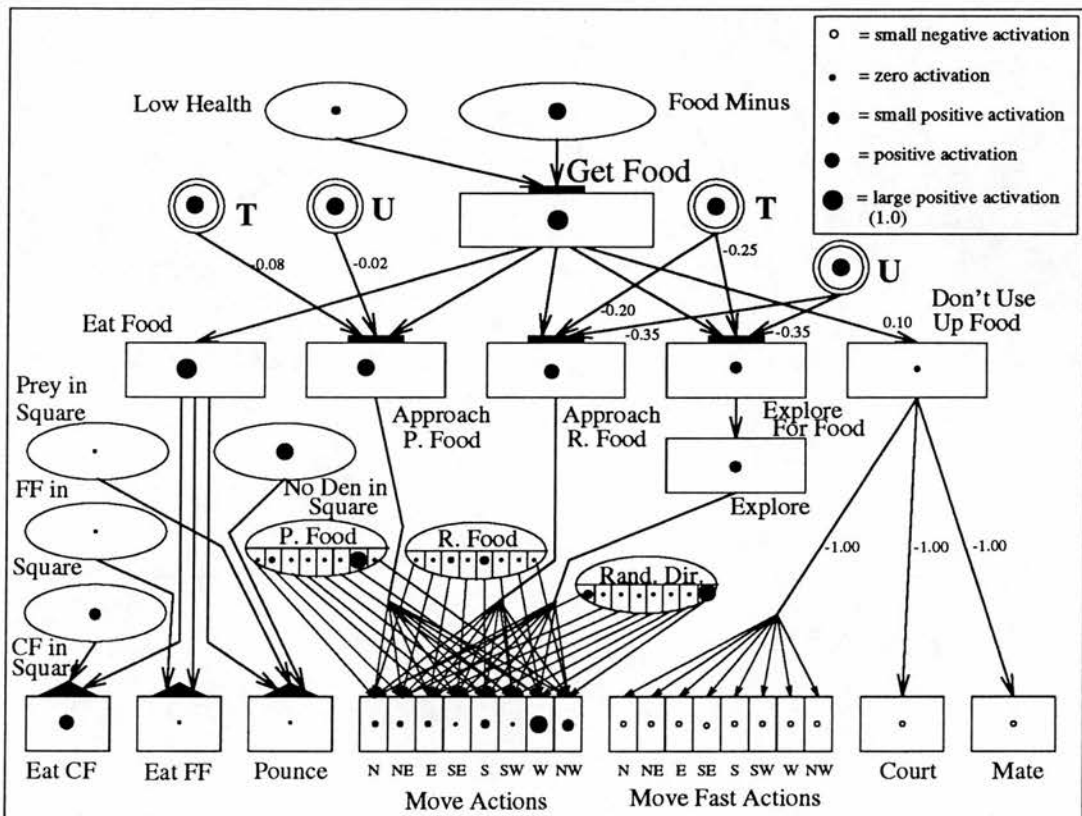


Figure 11.12: An example of the passage of activation through the 'get food' system. The possible effects of other systems on the action node activations are not considered. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11.

There are no major inadequacies in the performance of this ASM and it satisfies all of the criteria of section 10.5, with a couple of partial exceptions as explained below. However, this is only a preliminary, unfinished study and there are several aspects of the extended Rosenblatt & Payton model which are worthy of further investigation:

1. **persistence** – as noted in section 10.6.4, this ASM could perhaps benefit from some increased tendency towards persistence and contiguous action sequences. The fact that distance inversely affects the size of a stimulus tends to enforce

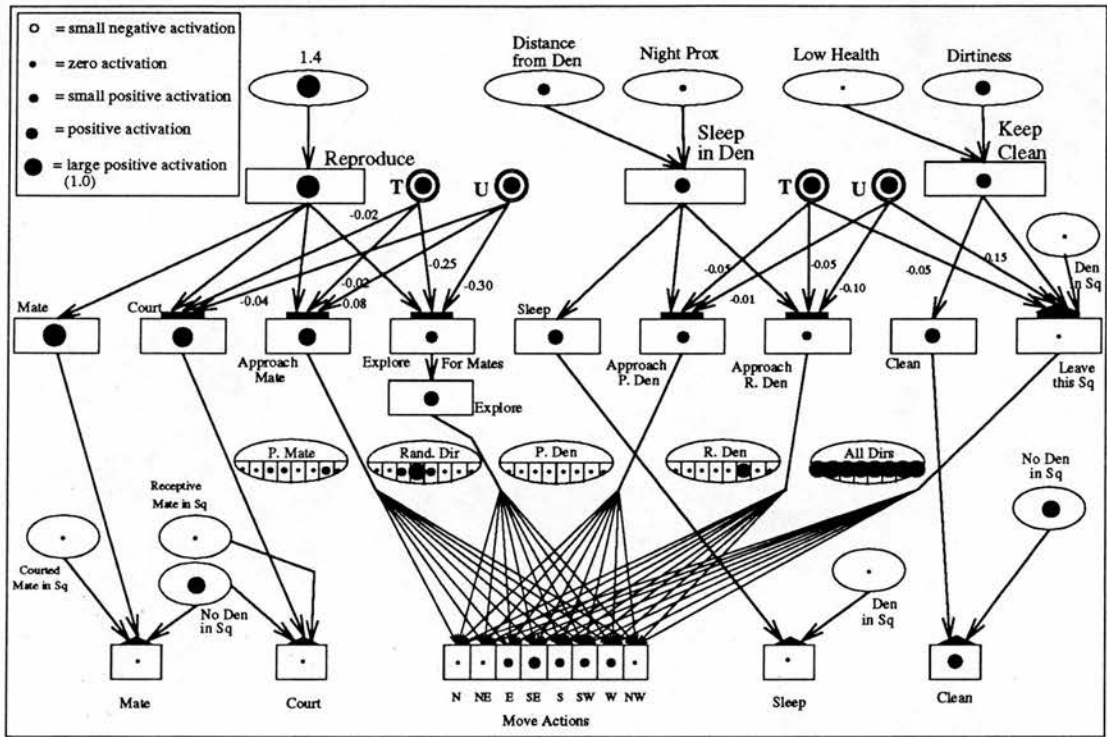


Figure 11.13: An example of the passage of activation through the 'reproduce', 'sleep in den' and 'keep clean' systems. The possible effects of other systems on the action node activations are not considered. An explanation of the meaning of some of the symbols in the diagram can be found in figure 9.11. The calculation of the activation of the node 'leave this square' is unusual: the input from 'keep clean' is added to the penalty inputs before being multiplied by the input from 'den in square'. The connections from the five multi-directional external stimuli to the MOVE actions have been omitted to avoid cluttering up the diagram.

persistence to some extent, in that a resource (e.g. food, water, a mate) which is at hand will have a larger stimulus value than one which is further away (assuming that the resources have equal intrinsic values). Beyond the indirect influence of stimulus values which decrease with distance though, there is nothing in the ASM to promote persistence. It may prove to be the case that changing the mechanism somehow to increase persistence will give rise to better performance.

2. **contiguous action sequences** – this is a similar issue to persistence. Again, the decreasing value of stimuli with distance tends to promote contiguous action sequences, at least in those appetitive-consummatory sequences which involve exploring and approaching before the consummatory act. The mechanism may benefit from explicit additions to promote the continuation of sequences once started.
3. **combination of preferences** – the present method of combining preferences is something of a fudge. Further study may lead to more principled ways of combining preferences, perhaps with functions that are tailored to the individual nodes rather than general-purpose rules.
4. **temporal and uncertainty penalties** – the values of these are at present somewhat arbitrary. It may be possible, with further study, to develop a more principled way of determining how large these penalties should be in each case.

In addition to these four points, there are probably other features that can be added to the ASM to improve its performance. The list of criteria in section 10.5 is almost certainly incomplete. Better understanding of the computational nature of the action selection problem will probably lead to further ways of improving the ASM proposed here.

One alternative to both hierarchical decision structures and free-flow hierarchies is Halperin's concept of *partial competition* (see section 8.7), in which there is a weak competitive inhibition between nodes at the same level, rather than a winner-take-all



inhibitory competition or no inhibition at all. Halperin's mechanism has not been tested here, but it is possible that there would be problems with it due to loss or distortion of information, and the arguments about the sensory bottleneck would apply in this case as well.

When discussing the performance and validity of her own mechanism [Maes, 91a], Maes talked about the need for a mechanism to be distributed and robust and to be both reactive and "motivation-oriented". The extended Rosenblatt & Payton mechanism proposed in this section is also distributed and robust. There are no overridingly crucial components in the extended Rosenblatt & Payton ASM and the only connections between systems are those to shared action and sub-system nodes. There will need to be some sort of machinery to implement the winner-take-all process between the different actions, but this is required of any sort of mechanism. Likewise, there are no global controlling structures and each system is to a large extent independent. The mechanism is also reactive: it can respond immediately and flexibly to changes in the environment. It does not follow any plans formulated in the past, and it takes input from stimuli describing only the current state of the internal and external environment. At the same time it is goal-driven: large disturbances in internal variables or other urgent situations will tend to lead to what appears from the outside to be purposive behaviour. The behaviour will tend to be organised so as to reduce the offset of a homeostatic variable or to remedy the urgent situation.

Even if any of the basic aspects of this ASM prove to be a bad idea, which seems unlikely, there are still things that can be learned from the way in which the mechanism was derived. The theoretical arguments presented here to motivate the design of the ASM are useful in their own right. It is also hoped that the overall methodology used here will prove instructive to others wanting to postulate different ASMs. The process of providing a solid theoretical basis for the mechanism, specifying the mechanism unambiguously, implementing and testing it in some sort of rigorous and complex simulation, and taking account of any shortcomings that are exposed therein is a sensible way of producing a viable and theoretically interesting mechanism.



Dawkins [1976] puts forwards hierarchies as a "candidate principle for ethology". He suggests that hierarchies are to be preferred because they are "big enough for the job" and because they provide "a way of making complexity manageable" [Dawkins, 76, pp 48]. The arguments put forward here for preferring free-flow hierarchies of a particular sort are in general more specific than those used by Dawkins, but it is also worth noting that the extended Rosenblatt & Payton-type hierarchy is the only type of structural organisation considered here that seems able to deal sensibly with all of the many facets of the complex problem of action selection. It is flexible enough to be able to handle internal, indeterminate and external stimuli, as well as consummatory and appetitive alternatives within a system and alternatives that differ in other ways. It is also able to produce appropriate behaviour for homeostatic and non-homeostatic systems, urgent and non-urgent systems, and prescriptive and proscriptive systems. It allows for arbitrary combination of stimuli and combination of preferences. The problem of action selection is a complex and intricate one, but something along the lines of the extended Rosenblatt & Payton mechanism discussed here seems able to deal with it.

## Chapter 12

# Conclusions to Part II

This second part of the thesis has covered the use of the simulated environment for testing and evaluating action selection mechanisms. Several action selection mechanisms were tested, and predicted and unpredicted difficulties were identified. Theories were presented to explain why these shortcomings in performance occurred and also to explain more subtle differences in performance between the tested action selection mechanisms. In section 10.5 a list was drawn up, the elements of which can be thought of in two ways; they can be thought of as both shortcomings in existing action selection mechanisms and as essential criteria to be taken into account in the design of future action selection mechanisms. Section 10.6 briefly enumerates the deficiencies that were uncovered while testing the different mechanisms (drives, Lorenz, Maes, and Rosenblatt & Payton) in the simulated environment. In chapter 11, a particular approach to action selection, that of Rosenblatt & Payton [1989], with various extensions by the author, was analysed more fully, and was suggested as the best approach for computing action selection.

One reason why there are so many computational inadequacies with the different mechanisms is that there has in the past been a lack of means of testing or validation for models of action selection / behavioural choice. Models have been proposed to some

extent in a vacuum, with the only means of proving or falsifying them lying in 'thought experiments'. Since people have a limited capacity to conceptualise the interactions between complex systems (in this case between a complex simulated environment and a complex action selection mechanism), many faults with mechanisms have gone undetected. To give two instances from this thesis: the difficulties with division rules in Maes' spreading activation network and the problems with straightforward combination (simple summation) of stimuli in the Rosenblatt & Payton mechanism were unsuspected beforehand and were only detected through the use of the simulated environment.

As well as providing some insights into the inadequacies of different schemes for selecting actions, the work in this second half of the thesis has also contributed to an understanding of the computational nature of the problem of action selection. Some of the contributions were: *(i)* understanding how the abstract need to maximise the future expected genetic fitness can be translated into more concrete and immediate criteria (section 9.2), *(ii)* understanding that the different sub-problems (such as avoiding hazards, getting food, not getting lost, reproduction) differ in many ways, and what some of these differences are (section 7.2), *(iii)* understanding the need for positive and negative preferences to be combined, and the difficulties involved in doing this (sections 11.1.2 and 11.1.3), and *(iv)* understanding that different alternatives (e.g. consummatory and appetitive), in the same system and in different systems, need to be arbitrated between in a balanced way (see sections 9.3.3 and 11.1.4).

Understanding the problem is a large part of finding the answer and due consideration has been given here to the exact nature of the problem of action selection for animals.

## 12.1 Further Work

Some suggestions for further work follow:

1. More action selection mechanisms (e.g. Halperin's mechanism as described in section 8.7) should be implemented and tested in the simulated environment described here.
2. The mechanism described in chapter 11 could be used for decision-making in other areas in which decisions have to be made while taking into account multiple goals of varying importance. Only the application to action selection for animals has been discussed here.
3. The suggested method for computing action selection presented here has only been tested in this simulated environment. It would be interesting to see how it performs in other simulated environments and in controlling the behaviour of a robot with real sensor input and real motor effectors.
4. In addition to trying to produce optimal action selection mechanisms, the simulated environment could be used to try and investigate the utility or 'purpose' of observed phenomena in animals such as time-sharing, displacement behaviour and vacuum activity.

# Epilogue

The two parts to the thesis have included four important contributions to the study of action selection:

1. explanation and description of a complex and fairly realistic simulated environment which models the problem of action selection or behavioural choice facing animals in the wild.
2. evaluation and analysis of several different action selection mechanisms using this simulated environment, and the discovery of inadequacies in the mechanisms.
3. enumeration and explanation of fourteen theoretical requirements needing to be incorporated into a capable action selection mechanism.
4. explanation and analysis of the ideas behind a mechanism which, it is suggested, provides a means for computing action selection in an optimal or near-optimal manner.

The initial idea of using a simulated environment has been shown to be successful by the results obtained. In particular, some of the problems with Maes' mechanism and the drawback to simple combination of preferences in Rosenblatt & Payton's mechanism are not at all obvious and would not have been detected without the explicit implementation and testing made possible by the use of the simulated environment.

The process of implementation and subsequent observation of performance, even in an imperfect and to some extents simplified simulated environment, is invaluable because it brings to the fore any ambiguities in the specification and any gross deficits in the performance of the mechanism which might not be predictable from introspection alone. Part of the reason for the quite significant errors found in some mechanisms is that no rigorous and realistically complex simulated environment, or any other comparable means of evaluation, has been available to test models or mechanisms in the past. The simulated environment described here has been shown to be a valid and useful investigatory and exploratory tool.

## **Appendix A:**

# **Details of the Implementation of the Extended Rosenblatt & Payton Mechanism**

This appendix describes the extended Rosenblatt & Payton mechanism in detail, as it was implemented to tackle the action selection problem embodied in the SE. The stimuli, nodes and connections of every system of the ASM are shown in a set of diagrams (figures A.1 – A.15), one for each system.

Shaded ellipses in the diagrams denote stimuli, some of which are multi-directional (i.e. have a component for each compass direction). Shaded rectangles denote nodes. Every connection has a strength or weight of 1.0 unless otherwise indicated.

Some symbols that are used in the figures are explained in figure 9.11. Strict summation of the inputs to a node is indicated by a solid rectangle, and strict multiplication by a solid triangular shape. For nodes which receive more than one input, but which have no rectangle or triangle, then the activation is in most cases calculated according to the rule described in section 11.1.3. The combination rules for these nodes are all given at the end of this appendix.

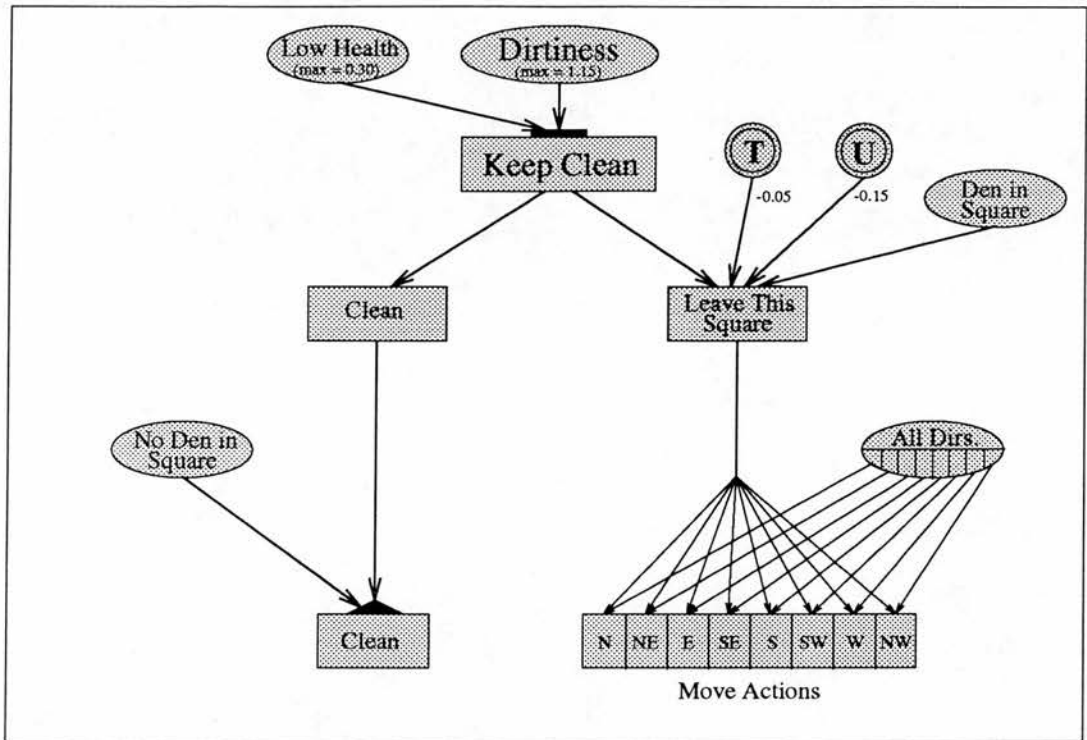


Figure A.1: Diagram showing implementation of 'keep clean' system. For further explanation see text and figure 9.11.

For all stimuli with names like 'DP' or 'Edges' then the name implies perception of the named type of feature. 'P' stands for perceived ('P. Food' is perceived food) and 'R' stands for remembered.

## Notes

1. **keep clean:** the animal is not allowed to clean/preen/groom itself when it is in its den.



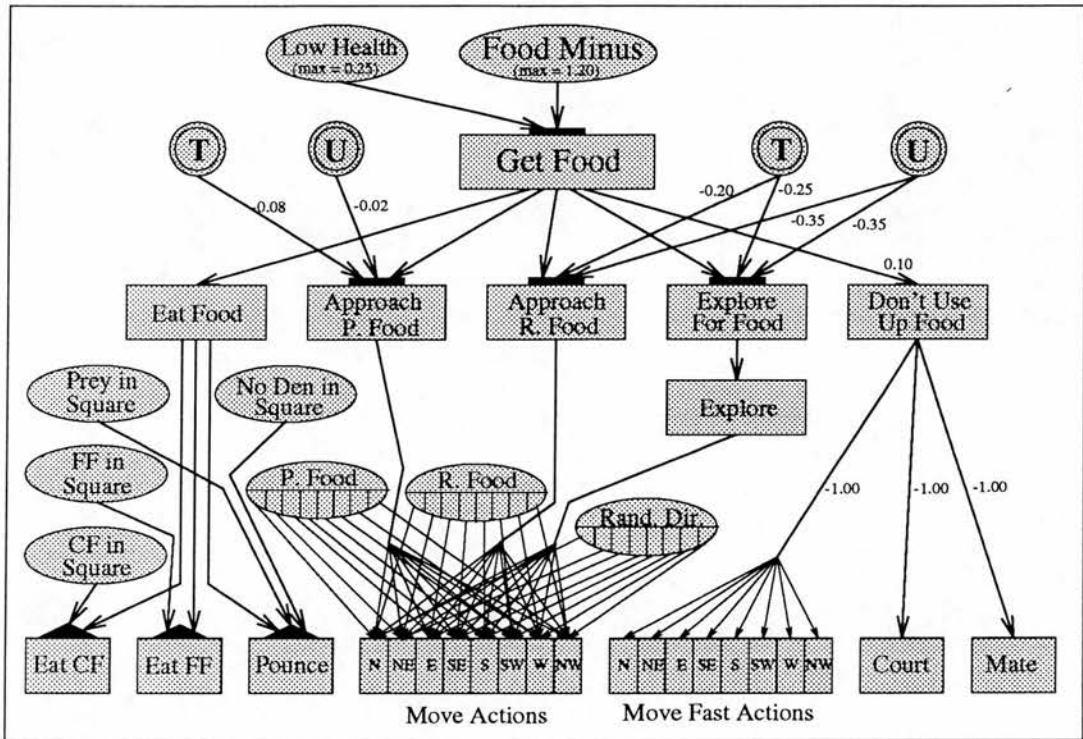


Figure A.2: Diagram showing implementation of 'get food' system. For further explanation see text and figure 9.11.

2. **get food:** the animal is not allowed to perform the action POUNCE while in its den. Various strenuous actions which will increase the animal's food deficit are inhibited to a small degree.
3. **get water:** Various strenuous actions which will increase the animal's water deficit are inhibited to a small degree.
4. **keep cool:** the two inputs to the system-level node are as follows: (i) *int\_temp\_plus* is a measure of how much the animal's actual body temperature is too high (i.e. the positive offset, if there is one, from the homeostatic set-point), and (ii) *ext\_temp\_plus* is a measure of how much the animal's body temperature would be too high if it were performing a very strenuous action out in the open (i.e.

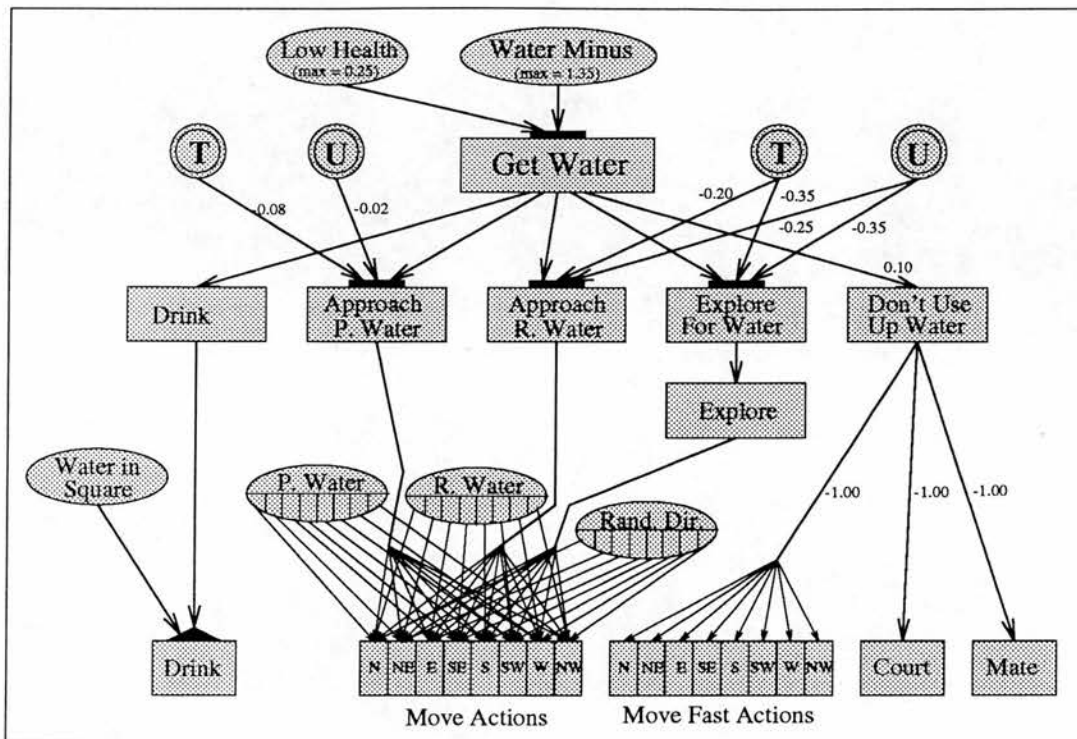


Figure A.3: Diagram showing implementation of 'get water' system. For further explanation see text and figure 9.11.

the potential worst-case scenario given the current temperature in the environment). If the system-level node receives a lot of excitation then there will be a mild inhibition of those strenuous actions which would increase the animal's temperature.

5. **keep warm:** the two inputs to the system-level node are as follows: (i) *int\_temp\_minus* is a measure of how much the animal's actual body temperature is too low (i.e. the negative offset, if there is one, from the homeostatic set-point), and (ii) *ext\_temp\_minus* is a measure of how much the animal's body temperature would be too low if it were performing a non-energetic action out in the open (i.e. the potential worst-case scenario given the current temperature

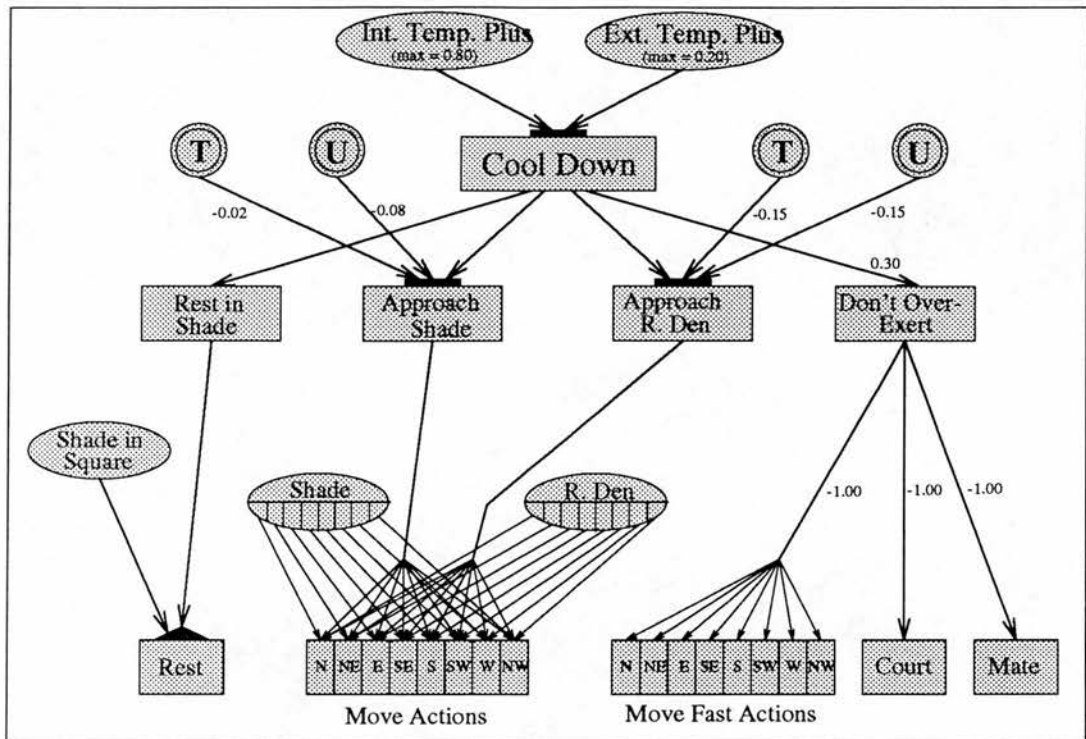


Figure A.4: Diagram showing implementation of 'keep cool' system. For further explanation see text and figure 9.11.

in the environment). If the system-level node receives a lot of excitation then there will be a mild excitation of those strenuous actions which would increase the animal's temperature.

6. **avoid DPs:** this system is not totally proscriptive because of the very fallible perception of the animal. A dangerous place / hazard may be present in an adjacent square to the animal and the animal may well not perceive it. Because of the noisy and error-prone perception and the very severe injuries that can be caused by hazards, it is better for the animal to stay away from hazards and not to approach them too closely. In this diagram (figure A.6) the different stimuli are:

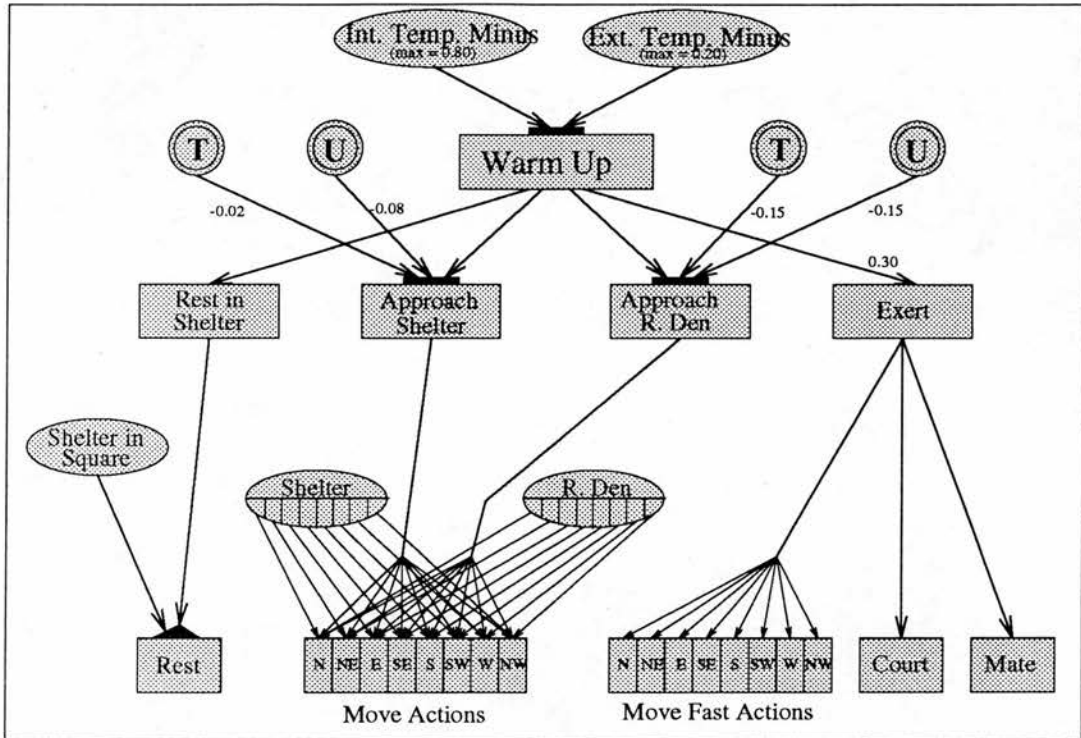


Figure A.5: Diagram showing implementation of 'keep warm' system. For further explanation see text and figure 9.11.

'Opp DP' – zero in the direction of the hazard and adjacent directions, 1.0 in the exact opposite direction and then slightly less than 1.0 in the directions adjacent to the opposite one.

'DP(1)' – perception of hazards / dangerous places in adjacent squares.

'DP(2)' – perception of hazards / dangerous places in squares that are two away from the animal's square.

7. **avoid irrs** : this system is again not totally proscriptive because of the very fallible perception of the animal. An irrelevant animal may be present in an adjacent square to the animal and the animal may well not perceive it. Because of

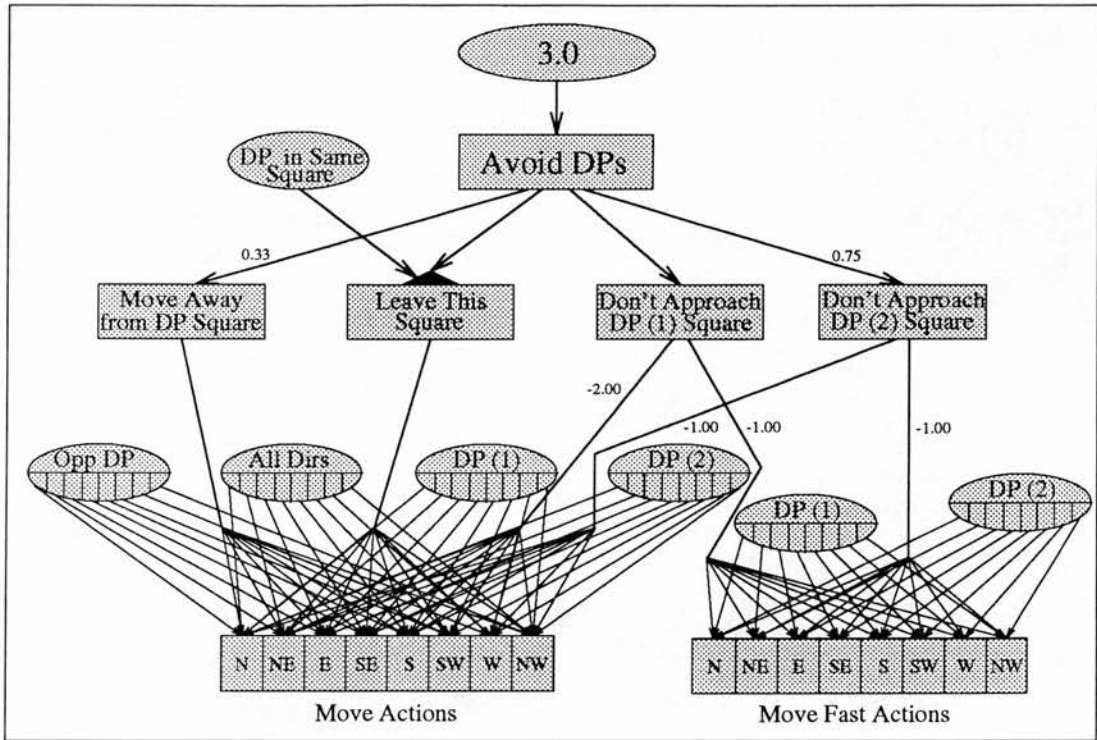


Figure A.6: Diagram showing implementation of 'avoid DPs' system. For further explanation see text and figure 9.11.

the noisy and error-prone perception and the severe injuries that can be caused by irrelevant animals, it is better for the animal to stay away from them and not to approach them too closely. In this diagram (figure A.7) the different stimuli are:

'Opp Irr' – zero in the direction of the irrelevant animal and adjacent directions, 1.0 in the exact opposite direction and then slightly less than 1.0 in the directions adjacent to the opposite one.

'DP(1)' – perception of irrelevant animals in adjacent squares.

'DP(2)' – perception of irrelevant animals in squares that are two away from the animal's square.

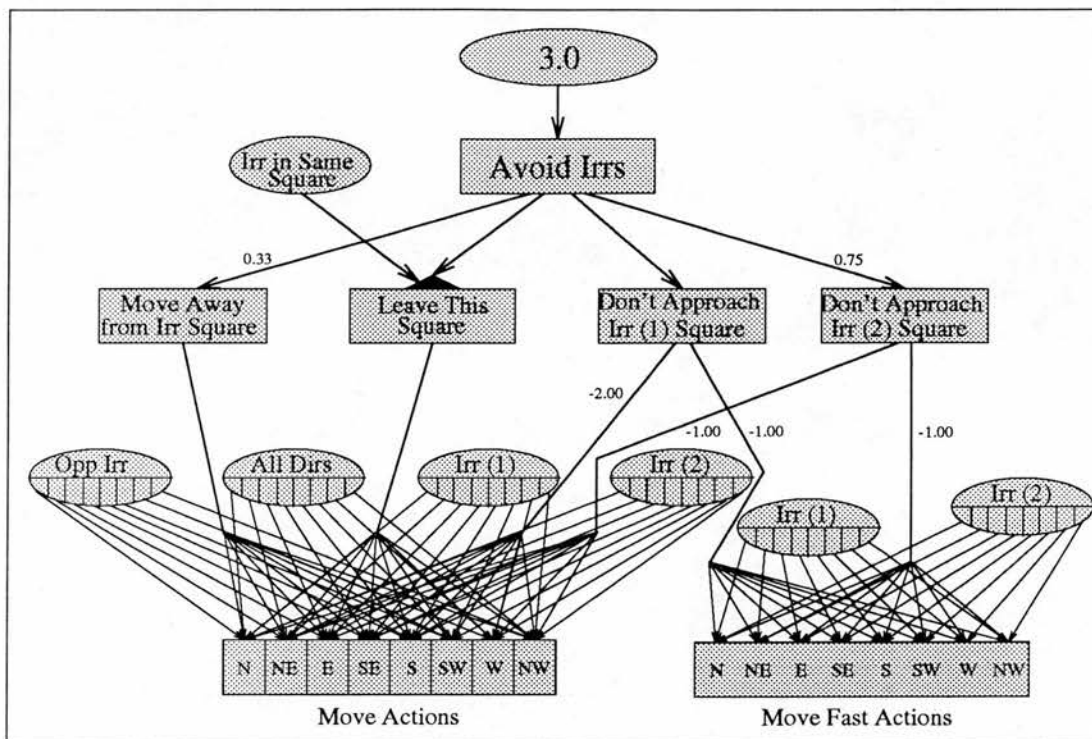


Figure A.7: Diagram showing implementation of 'avoid irrs' system. For further explanation see text and figure 9.11.

8. **avoid P1s:** the different stimuli in this system are:

'Max P1' – the maximum value, over the 8 different directions, of the perceptions of predator1s.

'Max P1 (adj)' – 0.0 if 'Max P1' is less than 0.30, otherwise equal to 'Max P1'. This is required to stop the animal choosing the action FREEZE too frequently when it is in dense vegetation (which produces a large value of 'shelter in square' and also reduces the efficacy of perception, producing many incorrect, low-valued perceptions of predator1s).

'Opp P1' – zero in the direction of the predator1s and adjacent directions, 1.0 in the exact opposite direction and then slightly less than 1.0 in the directions

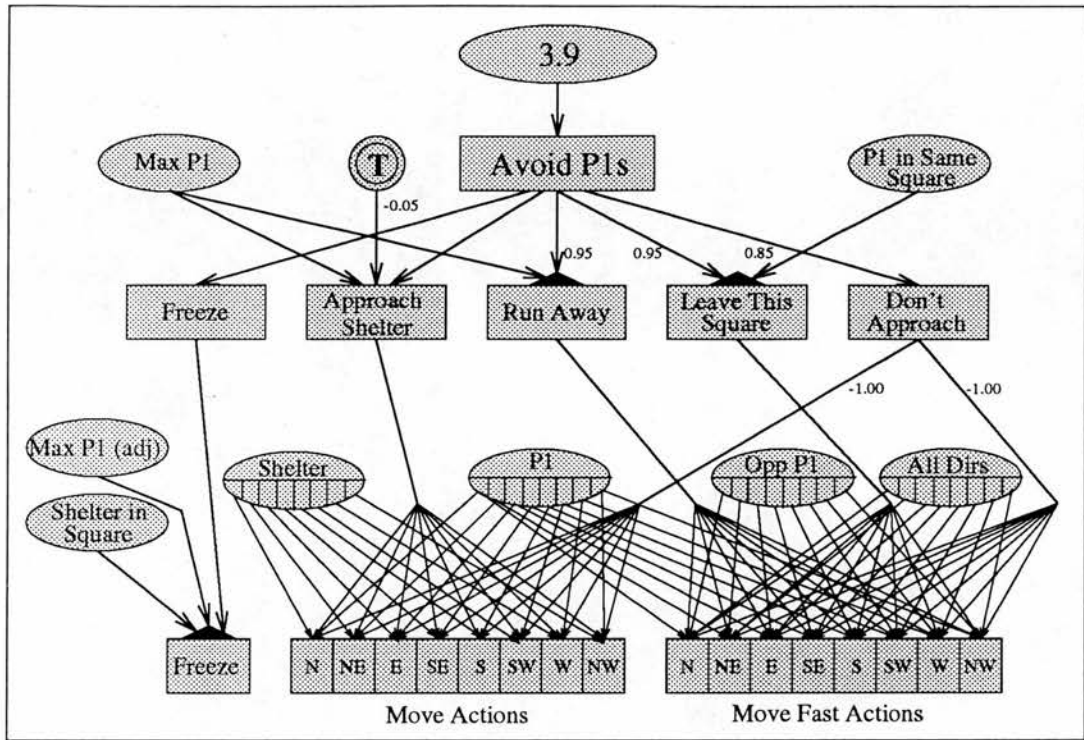


Figure A.8: Diagram showing implementation of 'avoid P1s' system. For further explanation see text and figure 9.11.

adjacent to the opposite one.

9. **avoid P2s:** the different stimuli in this system are:

'Max P2' – the maximum value, over the 8 different directions, of the perceptions of predator2s.

'Max P2 (adj)' – 0.0 if 'Max P2' is less than 0.30, otherwise equal to 'Max P2'. This is required to stop the animal choosing the action FREEZE too frequently when it is in dense vegetation (which produces a large value of 'shelter in square' and reduces the efficacy of perception, producing many incorrect, low-valued perceptions of predator2s).



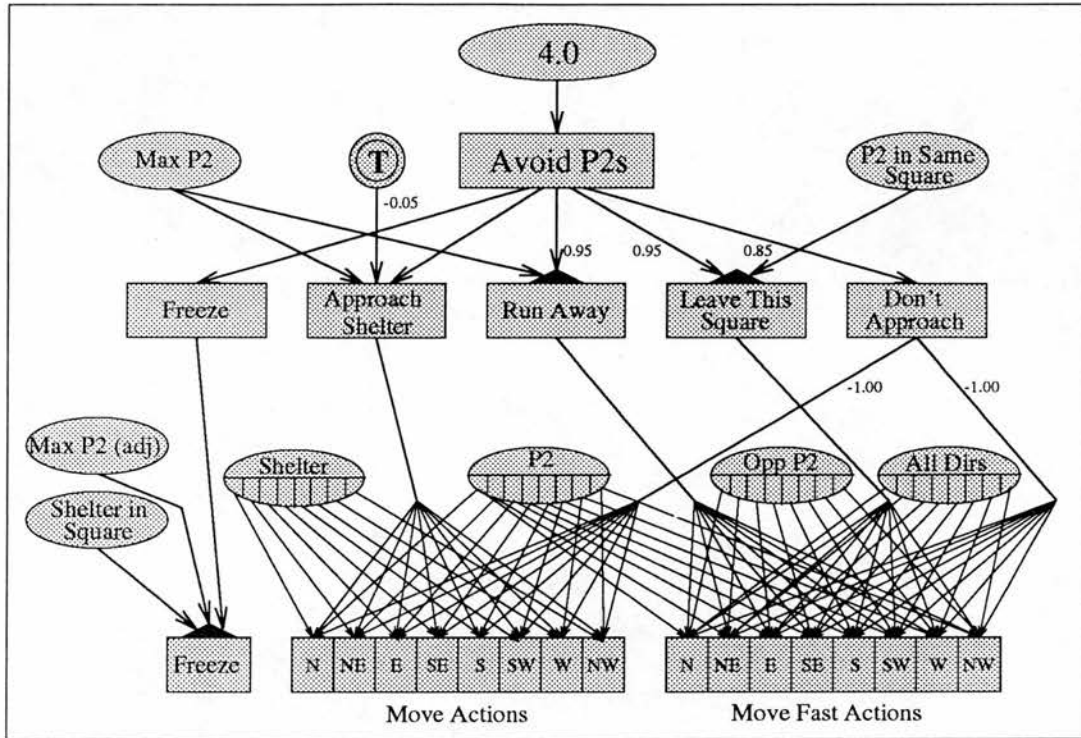


Figure A.9: Diagram showing implementation of 'avoid P2s' system. For further explanation see text and figure 9.11.

'Opp P2' – zero in the direction of the predator2s and adjacent directions, 1.0 in the exact opposite direction and then slightly less than 1.0 in the directions adjacent to the opposite one.

10. **reproduce:** the animal is not able to perform the actions COURT or MATE when it is in its den.
11. **sleep in den:** this system is crucial to the animal. If it does not manage to get to its den before nightfall, and stay there all night, then it stands a very large chance of dying during the night (see figure 4.10). This explains why the two stimuli to the system-level node have such large values. The two stimuli are combined in



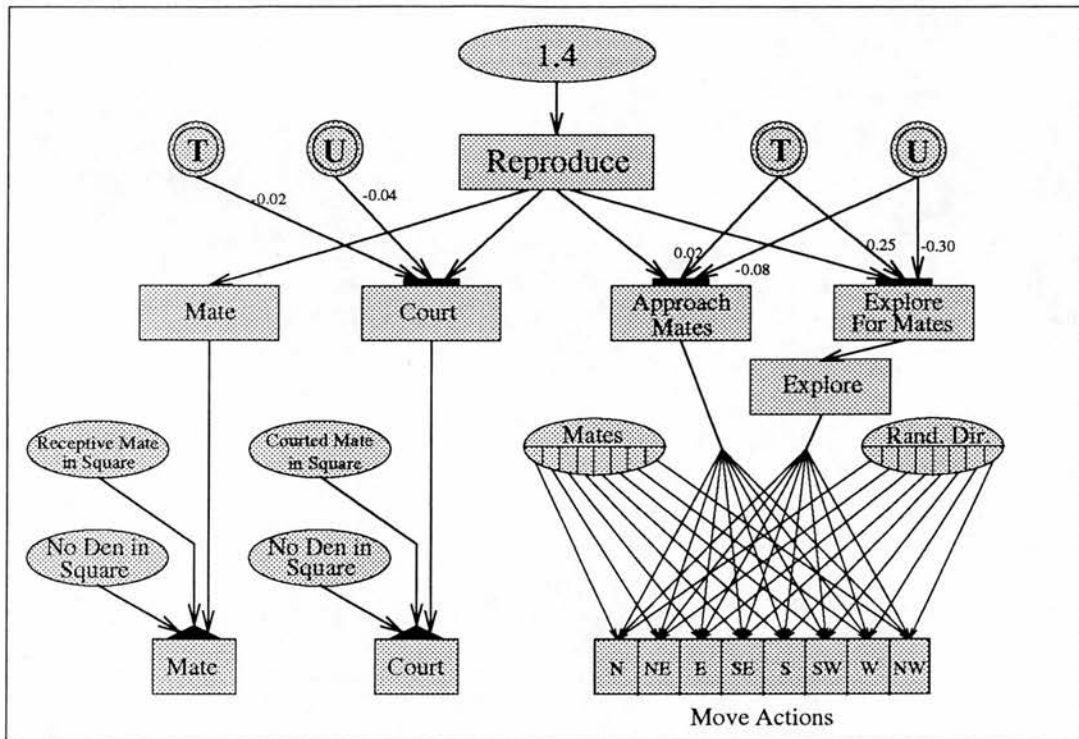


Figure A.10: Diagram showing implementation of 'reproduce' system. For further explanation see text and figure 9.11.

a slightly non-standard way, as described in the next section.

12. **keep variance low:** or 'do not get lost'. This is also a very important system because the animal is unlikely to be able to find its way back to its den before nightfall unless its estimated position with respect to the den is reasonably accurate. When the variance is quite high this system inhibits exploring, which tends to increase variance.
13. **scan for predators:** or 'vigilance'. The two stimuli 'P1 (adj)' and 'P2 (adj)' are transformed multi-directional perceptual inputs which have been altered so that they are equal to 1.0 if the same component of 'P1' is greater than 0.2, and are multiplied by  $\frac{1}{0.2}$  otherwise. This is because the animal needs to look in the

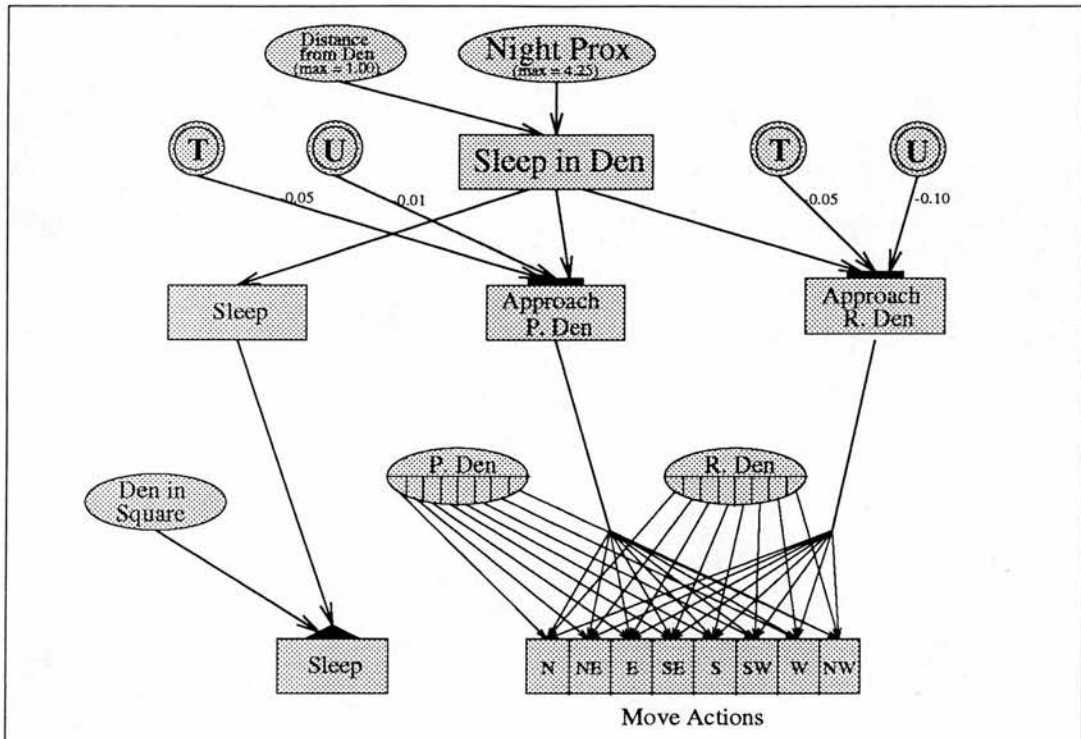


Figure A.11: Diagram showing implementation of 'sleep in den' system. For further explanation see text and figure 9.11.

direction of uncertain perceptions of predators.

14. **stay close to cover:** if the animal cannot perceive adequate cover then it is sensible for it to head back towards the den instead, where it will find cover eventually, if it does not encounter some on the way.

## Combination of Inputs

The activations for nodes which receive many inputs (and where the inputs are not just added or multiplied) are calculated in the following ways:

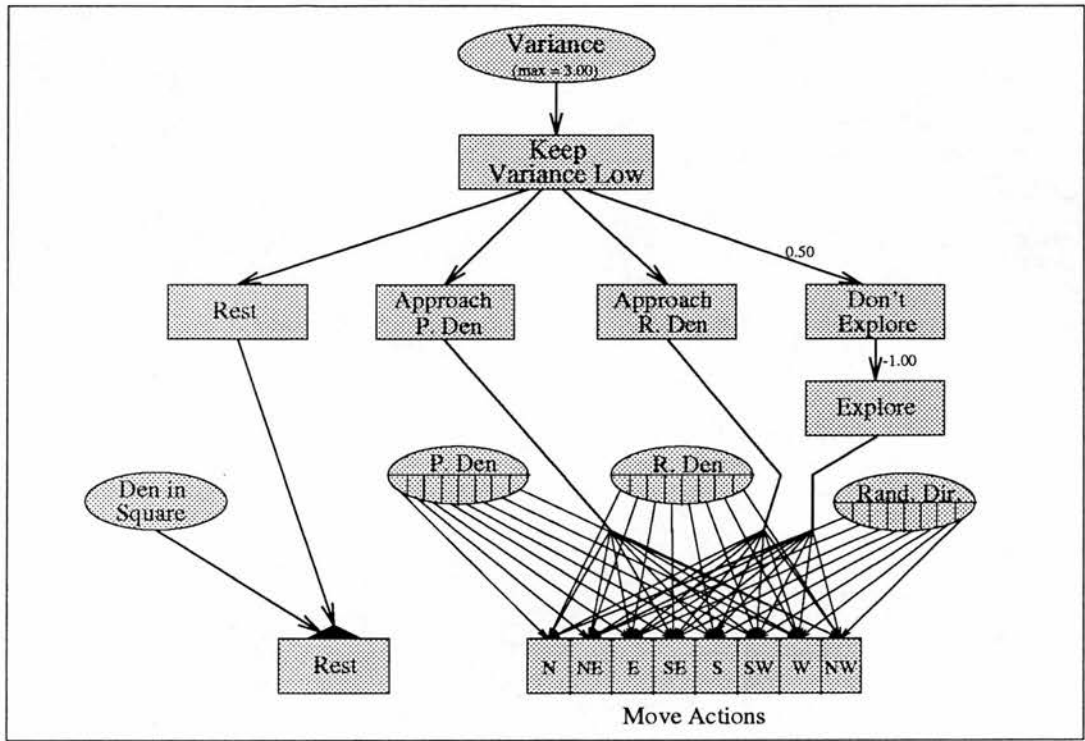


Figure A.12: Diagram showing implementation of 'keep variance low' system. For further explanation see text and figure 9.11.

$$A_j = S_1 + (S_1 \times S_2)$$

where  $A_j$  is the activation of the system-level node 'sleep in den',  $S_1$  is the indeterminate stimulus 'proximity of night' and  $S_2$  is the indeterminate stimulus 'distance from den'

$$A_j = (M \times E) + T$$

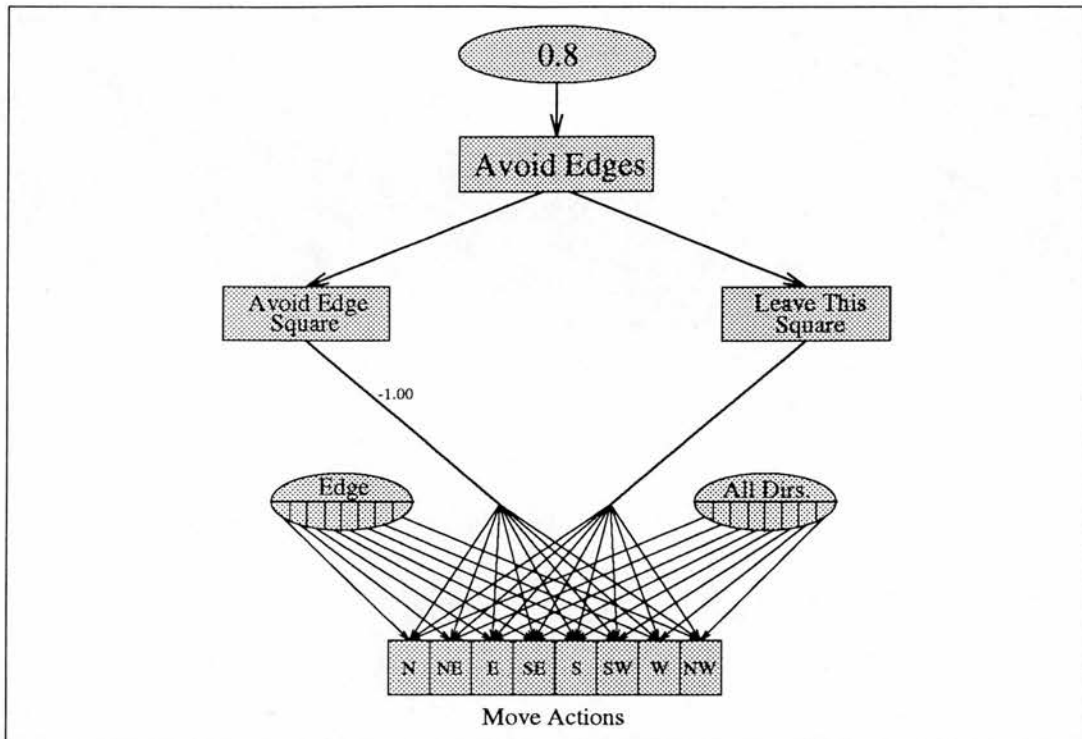


Figure A.13: Diagram showing implementation of 'avoid edges' system. For further explanation see text and figure 9.11.

where  $A_j$  is the activation of 'approach shelter' in the 'avoid predator1s' and 'avoid predator2s' systems,  $M$  is the (constant) motivational stimulus from 'avoid P1s' or 'avoid P2s',  $T$  is the (constant) negative penalty due to the fact that this is an appetitive action, and  $E$  is the external stimulus from 'max P1' or 'max P2'.

$$A_j = (M \times E) + T + U$$

where  $A_j$  is the activation of 'leave current square' in the 'keep clean' system,  $M$  is the motivational stimulus from 'keep clean',  $T$  and  $U$  are the negative penalties due to the

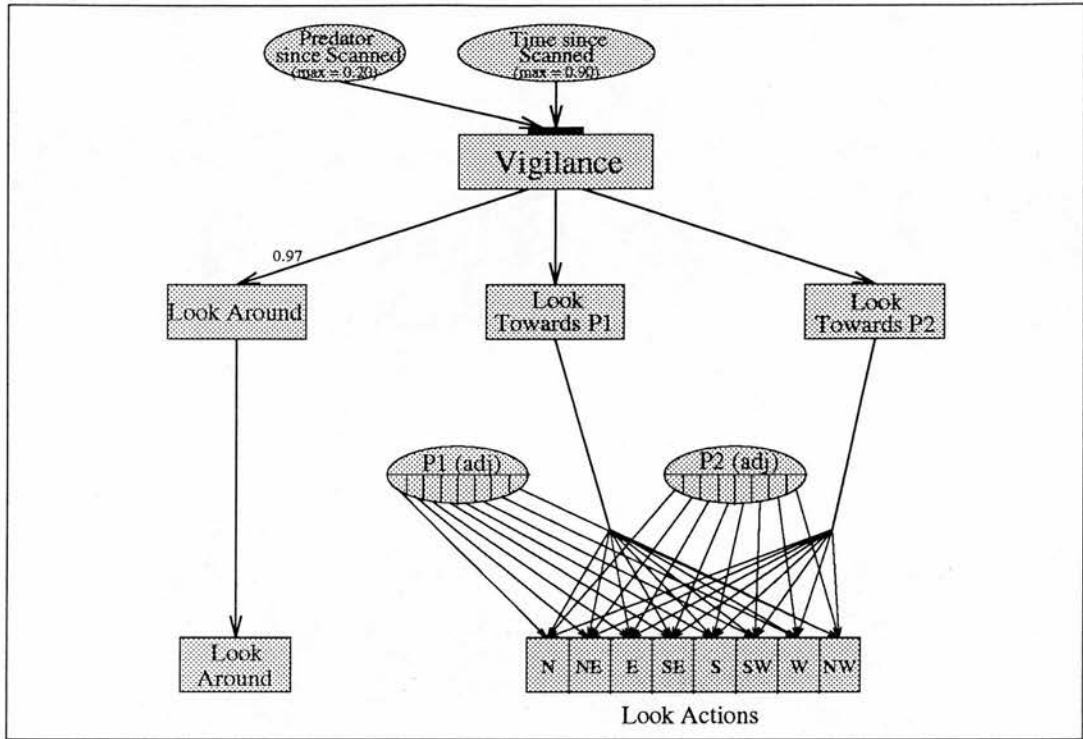


Figure A.14: Diagram showing implementation of 'scan for predators' system. For further explanation see text and figure 9.11.

fact that this is an appetitive action, and  $E$  is the external stimulus from 'den in square' (the animal is not able to clean itself while in its den).

$$A_j = \left( \frac{\max_i (P_i)_{i=1..3} + \frac{1}{9} \sum_{i=1}^3 (P_i)}{1 + \frac{1}{9}} \right) + P_4$$

where  $A_j$  is the activation of 'explore',  $P_1$ ,  $P_2$  and  $P_3$  are positive preferences for the node from the 'get food', 'get water' and 'reproduce', and  $P_4$  is a negative preference from the 'keep variance low' system.

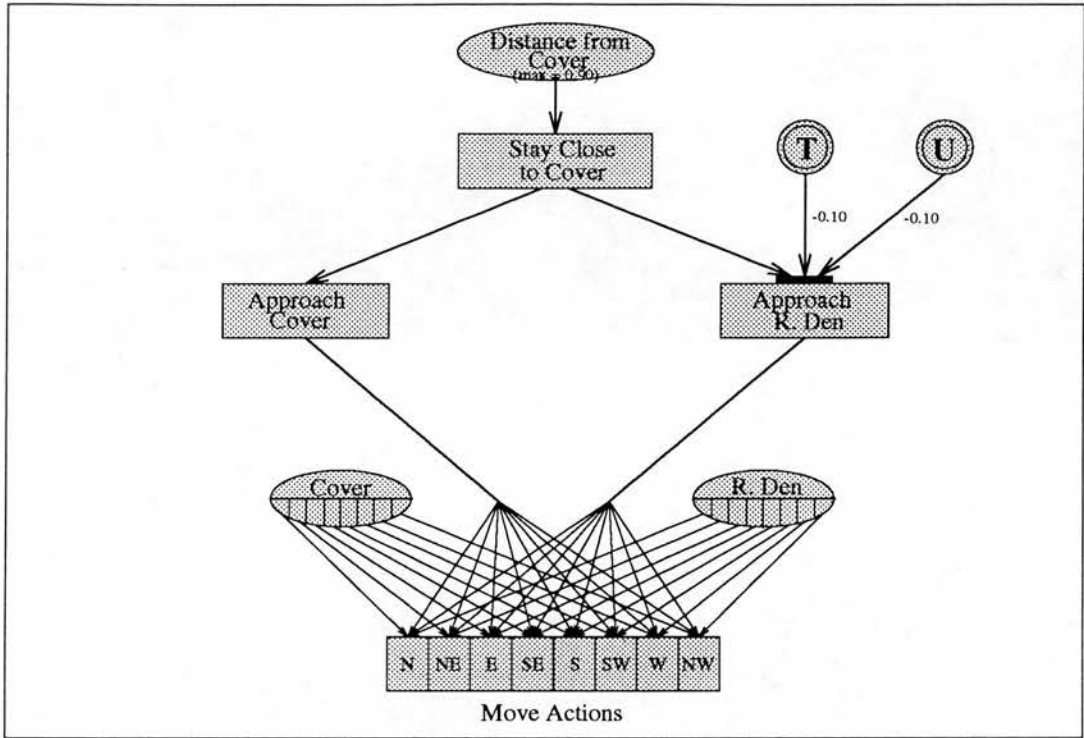


Figure A.15: Diagram showing implementation of 'stay close to cover' system. For further explanation see text and figure 9.11.

$$A_j = \left( \frac{\max_i(P_i)_{i=1...3} + \frac{1}{9} \sum_{i=1}^3 (P_i)}{1 + \frac{1}{9}} \right)$$

where  $A_j$  is the activation of action node 'rest',  $P_1$  is a positive preference obtained by multiplying a motivational stimulus from the 'keep cool' system with the external stimulus 'shelter in sq',  $P_2$  is a positive preference obtained by multiplying a motivational stimulus from the 'keep warm' system with the external stimulus 'shade in sq', and  $P_3$  is a positive preference obtained by multiplying a motivational stimulus from the 'keep variance low' system with the external stimulus 'den in sq'.

$$A_j = \left( \frac{\max_i(P_i)_{i=1,2} + \frac{1}{4} \sum_{i=1}^2(P_i)}{1 + \frac{1}{4}} \right)$$

where  $A_j$  is the activation of action node 'freeze',  $P_1$  is a positive preference obtained by multiplying a motivational stimulus from the 'avoid predator1s' system with both the external stimulus 'shelter in sq' and the external stimulus 'max p1 (adjusted)' and  $P_2$  is a positive preference obtained by multiplying a motivational stimulus from the 'avoid predator2s' system with both the external stimulus 'shelter in sq' and the external stimulus 'max p2 (adjusted)'.

$$A_j = \left( \frac{\max_i(P_i^+)_{i=1,2} + \frac{1}{4} \sum_{i=1}^2(P_i^+)}{1 + \frac{1}{4}} \right) + \left( \frac{\min_i(P_i^-)_{i=1...3} + \frac{1}{9} \sum_{i=1}^3(P_i^-)}{1 + \frac{1}{9}} \right)$$

where  $A_j$  is the activation of action node 'court',  $P_1^+$  is a positive preference obtained by multiplying a motivational stimulus from the 'reproduce' system with both the external stimulus 'receptive mate in sq' and the external stimulus 'den not in sq', and  $P_2^+$  is a positive preference from the 'keep warm' system.  $P_1^-$ ,  $P_2^-$  and  $P_3^-$  are negative preferences from the 'get food', 'get water' and 'keep cool' systems.

$$A_j = \left( \frac{\max_i(P_i^+)_{i=1,2} + \frac{1}{4} \sum_{i=1}^2(P_i^+)}{1 + \frac{1}{4}} \right) + \left( \frac{\min_i(P_i^-)_{i=1...3} + \frac{1}{9} \sum_{i=1}^3(P_i^-)}{1 + \frac{1}{9}} \right)$$

where  $A_j$  is the activation of action node 'mate',  $P_1^+$  is a positive preference obtained by multiplying a motivational stimulus from the 'reproduce' system with the external stimuli 'courted mate in sq' and 'den not in sq' and  $P_2^+$  is a positive preference from

the 'keep warm' system.  $P_1^-$ ,  $P_2^-$  and  $P_3^-$  are negative preferences from the 'get food', 'get water' and 'keep cool' systems.

$$A_j = \left( \frac{\max_i(P_i^+)_{i=1\dots 23} + \frac{1}{529} \sum_{i=1}^{23} (P_i^+)}{1 + \frac{1}{529}} \right) + \left( \frac{\min_i(P_i^-)_{i=1\dots 9} + \frac{1}{81} \sum_{i=1}^9 (P_i^-)}{1 + \frac{1}{81}} \right)$$

where  $A_j$  is the activation of one of the directional MOVE action nodes, each of  $P_1^+ - P_{23}^+$  are obtained by multiplying an input from higher in the hierarchy (e.g. from 'approach perceived food') with the relevant component of a multi-directional external stimulus (e.g. 'perceived food'), and each of  $P_1^- - P_9^-$  are obtained by multiplying an input from higher in the hierarchy (e.g. from 'avoid edge square') with the relevant component of a multi-directional external stimulus (e.g. 'perceived edge').

$$A_j = \left( \frac{\max_i(P_i^+)_{i=1\dots 5} + \frac{1}{25} \sum_{i=1}^5 (P_i^+)}{1 + \frac{1}{25}} \right) + \left( \frac{\min_i(P_i^-)_{i=1\dots 10} + \frac{1}{100} \sum_{i=1}^{10} (P_i^-)}{1 + \frac{1}{100}} \right)$$

where  $A_j$  is the activation of one of the directional MOVE.FAST action nodes, each of  $P_1^+ - P_4^+$  are obtained by multiplying an input from higher in the hierarchy (e.g. from 'run away' from predator1) with the relevant component of a multi-directional external stimulus (e.g. 'opposite P1'),  $P_5^+$  is a preference from the 'keep warm' system, each of  $P_1^- - P_7^-$  are obtained by multiplying an input from higher in the hierarchy (e.g. from 'don't approach predator1') with the relevant component of a multi-directional external stimulus (e.g. 'perceived predator1'), and finally  $P_8^-$ ,  $P_9^-$ , and  $P_{10}^-$  are negative preferences from the 'keep cool', 'get food' and 'get water' systems.



$$A_j = \left( \frac{\max_i (P_i)_{i=1,2} + \frac{1}{4} \sum_{i=1}^2 (P_i)}{1 + \frac{1}{4}} \right)$$

where  $A_j$  is the activation of one of the directional LOOK action nodes,  $P_1$  is a positive preference obtained by multiplying a motivational stimulus from the 'avoid predator1s' system with the relevant component of the multi-directional external stimulus 'p1 (adjusted)' and  $P_2$  is a positive preference obtained by multiplying a motivational stimulus from the 'avoid predator2s' system with the relevant component of the multi-directional external stimulus 'p2 (adjusted)'.

# Bibliography

- Ackley, D & Littman, M (1992). Learning from natural selection in an artificial environment. In: *Proceedings of the 2nd Artificial Life Conference*, (C Langton, R Taylor, J Farmer, & S Rasmussen, eds) , Addison Wesley.
- Agre, P (1990). Book review. *Artificial Intelligence*, **43**, 369–384.
- Agre, P & Chapman, D (1987). Pengi: An implementation of a theory of activity. In: *Proceedings of the Sixth National Conference on Artificial Intelligence [AAAI-87]* , Morgan Kaufmann.
- Albus, JS (1981). *Brains, Behaviours & Robotics*. BYTE Publications.
- Baerends, G (1976). The functional organisation of behaviour. *Animal Behaviour*, **24**, 726–735.
- Baerends, G & Drent, R (1970). The herring gull and its egg (Part I). *Behaviour Suppl XVIII*, , 265–310.
- Baerends, G & Drent, R (1982). The herring gull and its egg (Part II). *Behaviour*, **82**, 1–416.
- Barto, A, Sutton, R, & Anderson, C (1983). Neuron-like adaptive elements that can solve difficult learning control problems. In: *IEEE International Conference on Robotics and Automation* , IEEE.
- P Bateson & P Hopfer, eds (1985). *Mechanisms*, volume 6 of *Perspectives In Ethology*. Plenum Press.

- Beck, R (1983). *Motivation: Theories and Principles (2nd Edition)*. Prentice-Hall.
- Beer, R (1990). *Intelligence as Adaptive Behaviour: an Experiment in Computational Neuroethology*. Academic Press.
- Beer, R, Chiel, H, & Sterling, L (1990). A biological perspective on autonomous agent design. *Robotics and Autonomous Systems*, **6**, 169–186.
- Bolles, R (1975). *Theory of Motivation (2nd Edition)*. Harper and Row.
- Booker, L (1988). Classifier systems that learn internal world models. *Machine Learning*, **3**, 161–192.
- Braitenberg, V (1984). *Vehicles: Experiments in Synthetic Psychology*. MIT Press / Bradford Books.
- Brooks, R (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, **2**, 14–23.
- Brooks, R (1987a). Intelligence without representation. In: *Proceedings of the Workshop on the Foundations of Artificial Intelligence*, MIT Press.
- Brooks, R (1987b). A hardware retargetable distributed layered architecture for mobile robot control. In: *Proceedings of the IEEE Robotics and Automation Conference*, IEEE.
- Brooks, R (1990). Fast, cheap and out of control. *Science*, **248**, 959–961.
- Bruce, V & Green, P (1985). *Visual Perception: Physiology, Psychology and Ecology*. Lawrence Erlbaum Associates.
- Caruana, R & Schaffer, J (1988). Representation and hidden bias: Gray vs binary coding for genetic algorithms. In: *Proceedings of the Fifth International Conference on Machine Learning*, Morgan Kauffmann.
- Coderre, B (1989). Modelling behavior in Petworld. In: *Proceedings of the 1st Artificial Life Conference*, (C Langton, ed), Addison Wesley.
- Colgan, P (1989). *Animal Motivation*. Chapman Hall.

- Connell, J (1990). *Minimalist Mobile Robotics*. Academic Press Ltd.
- Davis, W (1979). Behavioural hierarchies. *Trends in Neuroscience*, **2**, 5–7.
- Davis, W, Mpitsos, G, & Pinneo, J (1974). The behavioural hierarchy of the mollusk pleurobronchea. *Journal of Comparative Physiology*, **90**, 207–243.
- Dawkins, R (1976). Hierarchical organisation: A candidate principle for ethology. In: *Growing Points in Ethology*, (P Bateson & R Hinde, eds). Cambridge University Press.
- Dawkins, R (1989). *The Selfish Gene (2nd edition)*. Oxford University Press.
- Dawkins, R & Carlisle, T (1976). Parental investment, male desertion and a fallacy. *Nature*, **262**, 131–132.
- Dewsbury, D (1978). *Comparitive Animal Behaviour*. McGraw Hill.
- Eibl-Eibesfeldt, I (1970). *Ethology: the Biology of Behaviour*. Holt, Rinehart and Winston, Inc.
- Ewart, J (1980). *Neuroethology – An Introduction to the Neurophysiological Fundamentals of Behaviour*. Springer-Verlag.
- Fentress, Je (1976). *Simpler Networks and Behaviour*. Sinauer Associates Inc.
- Firby, R (1987). An investigation into reactive planning in complex domains. In: *Proceedings of the Sixth National Conference on Artificial Intelligence [AAAI-87]*, Morgan Kaufmann.
- Fraenkel, G & Ginn, D (1940). *The Orientation of Animals, Kineses, Taxes and Compass Reactions*. Oxford University Press.
- Ginsberg, M (1989). Universal planning an (almost) universally bad idea. *AI Magazine*, **10**, 40–44.
- Goldberg, D (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley.

- Gould, J (1982). *Ethology: The Mechanisms and Evolution of Behaviour*. W.W. Norton & Co.
- Halliday, T (1983). Motivation. In: *Causes and Effects*, (T Halliday & P Slater, eds). Blackwell Scientific Publications.
- Halperin, J (1990). *Machine Motivation*. PhD thesis , University of Toronto.
- Halperin, J (1991). Machine motivation. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour*, (JA Meyer & SW Wilson, eds) , MIT Press / Bradford Books.
- Halperin, J (1992). Ethological connectionism: Synapses to strategies (unpublished report). Technical report.
- Hinde, R (1953). Appetitive behaviour, consumatory act and the hierarchical organization of behaviour – with special reference to the great tit (*Parus Major*). *Behaviour*, 5, 189–214.
- Hinde, R (1956). Ethological models and the concept of ‘drive’. *British Journal for the Philosophy of Science*, 6, 321–331.
- Hinde, R (1959). Unitary drives. *Animal Behaviour*, 7, 130–141.
- Hinde, R (1960). Energy models of motivation. *Symposium for the Society for Experimental Biology*, 14, 199–213.
- Hinde, R (1970). *Animal Behaviour: a Synthesis of Ethology and Comparative Psychology (2nd Edition)*. McGraw-Hill.
- Hinde, R (1982). *Ethology: Its Nature and Relations to Other Sciences*. Fontana Press.
- Houston, A (1982). Transitions and time-sharing. *Animal Behaviour*, 30, 615–625.
- Houston, A & McFarland, D (1980). Behavioural resilience and its relation to demand functions. In: *Limits to Action: The Allocation of Individual Behaviour*, (J Staddon, ed) pp. 177–203. Academic Press.

- Hull, C (1943). *Principles of Behaviour: an Introduction to Behaviour Theory*. D. Appleton-Century Company, Inc.
- Kandel, E (1979). *Behavioral Biology of Aplysia*. W.H. Freeman.
- Kilmer, W, McCulloch, W, & Blum, J (1969). A model of the vertebrate central command system. *International Journal of Man-Machine Studies*, **1**, 279–309.
- Koshland, D (1979). A model regulatory system: Bacterial chemotaxis. *Physiological Reviews*, **59**, 811–862.
- Kovac, M & Davis, W (1980). Neural mechanisms underlying behavioural choice in pleurobranchaea. *Journal of Neurophysiology*, **43**, 469–505.
- Kristan, W, Lockery, S, Wittenberg, G, & Cottrell, G (1989). Behavioural choice – in theory and in practice. In: *The Computing Neuron*, (R Durbin, C Miall, & G Mitchison, eds) pp. 180–204. Addison Wesley.
- Lea, S (1984). *Instinct, Environment and Behaviour*. Methuen.
- Lorenz, K (1950). The comparative method in studying innate behaviour patterns. *Symposia of the Society for Experimental Biology*, **4**, 221–268.
- Lorenz, K (1981). *Foundations of Ethology*. Springer-Verlag.
- Ludlow, A (1976). The behaviour of a model animal. *Behaviour*, **58**, 131–172.
- Ludlow, A (1980). The evolution and simulation of a decision maker. In: *Analysis of Motivational Processes*, (F Toates & T Halliday, eds). Academic Press.
- Maes, P (1989). How to do the right thing. *Connection Science*, **1**, 291–323.
- Maes, P (1990). Situated agents can have goals. *Journal for Robotics and Autonomous Systems*, **6**.
- Maes, P (1991a). A bottom-up mechanism for behaviour selection in an artificial creature. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour*, (JA Meyer & SW Wilson, eds) , MIT Press / Bradford Books.

- Maes, P (1991b). The agent network architecture (ANA). In: *Proceedings of the AAAI Spring Symposium on Integrated Intelligent Architectures*, AAAI Press.
- Maes, P (1991c). Adaptive action selection. In: *Programme of the Thirteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates.
- Maes, P & Brooks, R (1990). Learning to coordinate behaviours. In: *Proceedings of the AAAI Symposium* pp. 796–802, AAAI Press.
- Manning, A (1979). *An Introduction to Animal Behaviour (3rd Edition)*. Edward Arnold.
- Marler, P & Hamilton III, W (1966). *Mechanisms of Animal Behaviour*. Wiley.
- Marr, D (1982). *Vision*. W. H. Freeman & Co.
- McFarland, D (1969). Mechanisms of behavioural disinhibition. *Animal Behaviour*, **17**, 238–242.
- McFarland, D (1974a). Time-sharing as a behavioural phenomenon. In: *Advances in the Study of Behaviour*, (D Lehrman, J Rosenblatt, H R.A., & E Shaw, eds) volume 4. Academic Press.
- McFarland, D (1974b). *Motivational Control Systems Analysis*. Academic Press.
- McFarland, D (1977). Decision-making in animals. *Nature*, **269**, 15–21.
- McFarland, D (1985). *Animal Behaviour*. Longman.
- McFarland, D (1989). *Problems of Animal Behaviour*. Longman.
- McFarland, D & Houston, A (1981). *Quantitative Ethology: The State Space Approach*. Pitman.
- McFarland, D & Sibly, R (1972). Unitary drives revisited. *Animal Behaviour*, **20**, 548–563.
- McFarland, D & Sibly, R (1975). The behavioural final common path. *Philosophical Transactions of the Royal Society (Series B)*, **270**, 265–293.
- McNamara, J & Houston, A (1986). The common currency for behavioural decisions. *The American Naturalist*, **127**, 357–378.

- Moore, A (1990). *Efficient Memory-Based Learning for Robot Control*. PhD thesis University of Cambridge.
- Morris, D (1990). *Animal Watching*. Jonathan Cape.
- Mrosovsky, N (1990). *Rheostasis: The Physiology of Change*. Oxford University Press.
- Pebody, M (1991). *How to Make a Lego Vehicle Do the Right Thing*. Master's thesis Department of Artificial Intelligence, University of Edinburgh.
- Roitblat, H (1988). A cognitive action theory of learning. In: *Systems with Learning and Memory Abilities*, (J Delacour & J Levy, eds). Elsevier Science Publishers.
- Rosenblatt, K & Payton, D (1989). A fine-grained alternative to the subsumption architecture for mobile robot control. In: *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks*, IEEE.
- Sandefur, J (1990). *Discrete Dynamical Systems*. Oxford Clarendon Press.
- Silverston, A (1985). *Model Neural Networks and Behaviour*. Plenum Press.
- Sibly, R & McCleery, R (1985). Optimal decision rules for gulls. *Animal Behaviour*, **33**, 449–465.
- Sibly, R & McFarland, D (1974). State-space approach to motivation. In: *Motivational Control Systems Analysis*, (D McFarland, ed). Academic Press.
- Sibly, R & McFarland, D (1976). On the fitness of behaviour sequences. *American Naturalist*, **110**, 601–617.
- Sutton, R (1988). Learning to predict by the methods of temporal difference. *Machine Learning*, **3**, 9–44.
- Sutton, R & Barto, A (1987). A temporal-difference model of classical conditioning. In: *Proceedings of the Ninth Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates.



- Taylor, C, Jefferson, D, Turner, S, & Goldman, S (1989). RAM: Artificial life for the exploration of complex systems. In: *Proceedings of the 1st Artificial Life Conference*, (C Langton, ed) , Addison Wesley.
- Tinbergen, N (1950). The hierarchical organization of mechanisms underlying instinctive behaviour. *Experimental Biology*, **4**, 305–312.
- Tinbergen, N (1951). *The Study of Instinct*. Clarendon Press.
- Tinbergen, N (1963). On aims and methods of ethology. *Zeitschrift fur Tierpsychologie*, **20**, 410–433.
- Toates, F (1980). *Animal Behaviour – A Systems Approach*. Wiley.
- Toates, F (1986). *Motivational Systems*. Cambridge University Press.
- Travers, M (1989). Animal construction kits. In: *Proceedings of the 1st Artificial Life Conference*, (C Langton, ed) , Addison Wesley.
- Watkins, C (1989). *Learning from Delayed Rewards*. PhD thesis University of Cambridge.
- Wilson, S (1985). Knowledge growth in an artificial animal. In: *Proceedings of the First International Conference on Genetic Algorithms and their Applications* , Lawrence Erlbaum Associates.
- Wilson, S (1991). The animat path to AI. In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behaviour*, (J Meyer & S Wilson, eds) , MIT Press / Bradford Books.
- Wine, J (1984). The structural basis of an innate behaviour pattern. *Journal of Experimental Biology*, **122**, 283–319.