Possibly Optimal Decision-Making Under Self-sufficiency and Autonomy*

Emmet Spier[†‡] David McFarland[†]

Abstract

Any self-sufficient autonomous agent faces multiple tasks to which it must sequence its behaviour if it is to manage its time appropriately. This paper considers the minimal multiple task scenario - the two resource problem. Some possible mechanisms, inspired from ideas in ethology and having claim to perform optimally when analysed mathematically, are considered: one based on cost functions and another based on motivational tendencies. Several variants of the mechanisms are implemented and tested in a continuous two dimensional simulation environment under a variety of conditions. Those mechanisms closely linked to cost functions do not perform well, are not adaptive, and dither. The mechanisms based on motivational tendencies perform better, are more adaptive, and demonstrate opportunism. The difficulty of finding effective mechanisms by using a functional perspective is noted.

1 Introduction

Any autonomous agent needs to be able to resolve two things: what to do next and how to do it[1]. This paper concerns itself with the first of those two questions. In particular, we are interested in the class of self-sufficient autonomous agents. Self-sufficient agents must ensure they will never run out of fuel or place themselves in a situation that will lead to their running out of fuel (McFarland and Bösser, 1993).

An autonomous agent, much like an animal, must be able to sequence its behaviour between the various tasks that it must perform; at the very minimum these tasks would include maintaining its fuel supply and performing some additional tasks to, perhaps, earn its keep (McFarland and Bösser, 1993; McFarland and Spier, 1997). The requirement for an agent to be able to generate such a sequence is described under terms such as the action selection problem or behaviour sequencing. Sequencing of behaviour has received a great deal of interest and the problem has seen many approaches, for instance from Agre and Chapman (1987), Rosenblatt and Payton (1989) and Maes (1990) in the simulation of adaptive behaviour literature, from the more classical approaches in artificial intelligence (AI), as well as from some aspects of the foraging theory literature.

We agree with Brooks (1990) that the deliberative methods used in classical AI for action selection have, certainly, not provided a fully satisfactory solution to the action selection problem. Such methods require the construction and maintenance of large models of the world. As Wilson (1991) has argued, if the agent has been constructed with due regard for the environmental niche in which it will exist, then more reactive or stimulus-response methods of generating its behaviour sequences may be quite sufficient. The model of behaviour sequencing developed in this paper is such a reactive type and following Wilson (1991) we will call such artificial 'animals', animats.

The second constraint we have placed upon our agents is that they should be self-sufficient. By accepting this constraint on our agent we gain several benefits. Firstly, self-sufficiency gives the experimenter a concrete base line from which he can judge the performance of the agent; either the agent is self sufficient, or it is not. Within a multi-task scenario there is no reason to expect that the agent will ever finish any particular task, for instance window cleaning. As such, it can be difficult to assess how the agent has performed in an absolute sense but much easier to see if the agent is managing to prevent the size of its workload from continuously growing. Secondly, the self-sufficiency condition may give the agent a way of assessing its own behaviour, though the feedback from its own internal state, allowing the agent determine its own priorities without an externally imposed goal state. This can be useful for purposes of learning. Thirdly, the constraint of self-sufficiency has many parallels with the survival constraint in a biological context and, as such, it is tempting to draw upon the history of ideas from ethology (McFarland and Houston, 1981; Halperin, 1991; Tyrrell, 1993; Blumberg, 1994; Spier and McFarland, 1996).

One attractive idea that ethology has utilised for analysing the behaviour of animals is that of optimality, reviewed in Stephens and Krebs (1986). This is a

* *Journal of Theoretical Biology* in press.

[†]Animal Robotics Laboratory,
Animal Behaviour Research Group Zoology Department, Oxford University, South Parks Road, Oxford OX1 3PS, UK. emmet.spier, david.mcfarland@zoology.ox.ac.uk

[‡]Center for Computational Neuroscience and Robotics, University of Sussex, Brighton BN1 9RH, UK. emmet@cogs.sussex.ac.uk

[1]It is not necessarily the case that the answer to one of these questions cannot provide a guide to the other, however, Spier and McFarland (1996) show a possible avenue for development.

functional criterion. For an agent, the concept of optimality must be defined with respect to a specification of performance which constitutes the function of the agent. If we could design agents that behaved optimally given the constraints we place upon them, such as achieving tasks we wish to see completed, then we cannot expect to obtain better results from that particular physical agent. Unlike animals, without the force of evolution, that statement is a tautology; however, optimality approaches still provide a very useful design guide since they provide methods to determine the upper bound to the performance of an agent where the concept of self-sufficiency provides the lower bound. There are two problems with the optimality approach.

The first is the requirement of finding an appropriate utility function. For a biological problem, this does not have to be difficult, the ultimate utility function for behavioural ecology is derived from the survival of the animal and its progeny. For more fine grain analysis, the utility function or currency can be guessed from this ultimate function and previously has been as simple as energy or more intricate like the quantity of plankton caught in a caddisfly net (Georgian and Wallace, 1981). McFarland and Bösser (1993) suggest possible 'ultimate' fitness functions for autonomous agents may be based upon satisfying the owner as opposed to just simply surviving. It should be noted, as Matarić and Cliff (1996) argue, the design of more fine-grained fitness functions that capture the agent-environment interactions can be a complex task.

The second, and perhaps more serious, problem when considering applying optimality to agent design is that there is no rôte method of obtaining a mechanism for generating behaviour if you are in possession of a utility function. For designers of agents this is unsatisfactory since knowledge of what the agent should achieve is not the same as how it should achieve it. Methods available for calculating behaviour from a given utility function often require a perfect model of the environment (Bellman, 1957). In some sense this approach is similar to the ones involved in classical planning where a goal state is required instead of a utility function. Typically, autonomous agents do not possess models of the environment, especially unpredictable dynamic environments (Georgeff, 1987). Alternatively a genetic algorithm (GA) approach can be taken (e.g. Cliff, Harvey and Husbands, 1993, have produced impressive results), however the mechanisms they generate are typically large recurrent neural networks and, as such, are inscrutable and atheoretical. Certainly GA methods are a plausible approach but, here, we are making an attempt to develop a more theoretical perspective upon decision-making.

This paper considers both the functional and mechanistic aspects involved in the design of a behaviour sequencing system for self-sufficient autonomous agents. In **2** we describe a functional framework, using optimality methods, within which we can consider decision-making and then develop a mechanistic decision-making model. In the subsequent sections we then test, within a spatial simulation environment, the mechanistic model against various interpretations of the functional model. This approach allows us to consider what we can learn from analytic methods as well as providing an attempt to extend our understanding of behaviour sequencing using reactive mechanisms.

2 The two-resource problem as a paradigm case for decision-making

A popular problem in animal psychology is the one-resource problem. This involves the study of animals controlling their behaviour to obtain some resource, for instance the control of feeding behaviour is reviewed by Staddon and Zanutto (1996), drinking by Rolls and Rolls (1982) and mating by McFarland and Nunez (1978). There are parallels, here, in the animat literature with energy as a resource (Wilson, 1985). From the perspective of investigating decision-making, these experiments are unsatisfactory because they do not offer the animal any possibility of trading-off between the various systems concerned with other resources. This presents a rather constrained paradigm for model making. Extending the investigation to two resources has the advantage that interactions between systems can be incorporated into any explanation of the behaviour. Examples are interactions of hunger and thirst, food and temperature, money and play (Toates, 1980, 1986; Cabanac, 1992).

It could be argued that three or more resources would impose extra, realistic, constraints upon a model of behaviour. This could be valid, but just considering two-resources adds a significant degree of extra complexity whereas adding a third does not make as much difference; the change from a scalar architecture to a vector one provides constraints upon the decision-making mechanisms that adding an another state to an already vector architecture would not. Additionally, it could be argued, that adding extra resources would not enlighten us any further because if achieving any particular resource is an exclusive act then we can always consider one resource to be competing against the bundle of possible attentions of all the other resources. An example of such a binary divide is energy, as the single resource, and work, as the collection of competing resources (McFarland and Spier, 1997). Of course, this perspective does not match the n-resource problem exactly but it does force us to consider our models with multiple resources in mind.

Within the animat field, work on the multiple resource problem has been considered by Blumberg (1994,

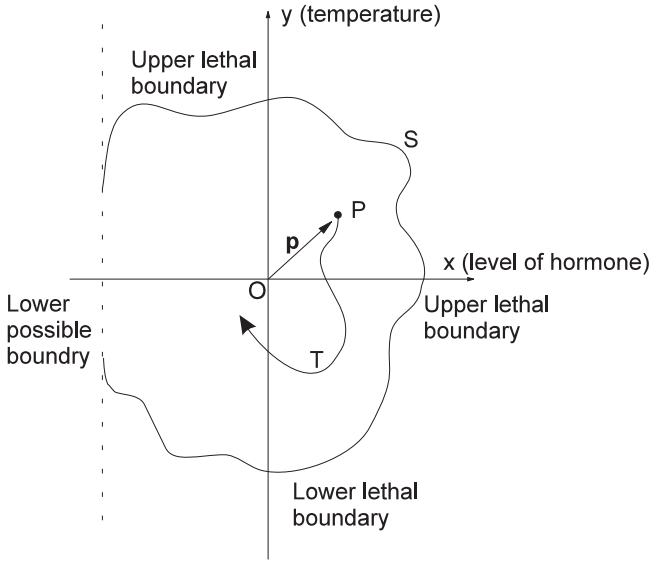

Figure 1: A state space. An example of a hypothetical two-dimensional physiological space with origin O corresponding to optimal body temperature and hormone level. The current physiological state is indicated by the point P or vector **p**. The boundary surface S separates the possible and lethal limits to values in the state space. T shows a possible trajectory the agent could take. After Sibly and McFarland (1974).

two resources) and Tyrrell (1993, many resources). They both consider the issues of satisfying several states however no criterion (except the ultimate one) was used to assess decision-making. Work discussed in **2.2** tries to provide a more analytic approach to assessing agent decision-making by considering that the agent can be attributed a cost function (its own or an 'owners'; McFarland and Bösser, 1993). The subsequent body of this paper investigates the developed theoretical results in a simulation environment.

2.1 The Agent as a state space

The state space model for agents was proposed by Sibly and McFarland (1974) and further developed in McFarland and Sibly (1975) and McFarland and Houston (1981). Within this framework an agent is characterised as a minimal set of internal state variables that can completely describe its physiological state. In such a description of an animal we could possibly identify hunger, thirst, temperature, oestrogen etc. as essential physiological state variables - for an animat we could identify energy, oil level, clean floors, etc. These variables sit within a Euclidean vector space as its orthogonal axes. In this space there will be regions that the agent can never encounter, for instance negative hormonal levels, and regions that should the agent cross into then it would die. The boundary of regions that are fatal to the agent are called lethal limits. Figure 1 illustrates such a state space for a model animal. The task of the agent within such a model is to maintain 'homeostasis' of its state variables under perturbations from its own behaviour (eating makes you thirsty) and the environment's impact on it (at night it gets colder). In terms of an animat we would assign an axis for each task that we wish it to achieve. This would form a minimal space in which we could describe any behaviour it makes as a vector in the state space of consequences.

2.2 The quadratic cost function model of the two-resource problem

Sibly and McFarland (1976) considered the decision-making process in the two-resource problem from a functional perspective by postulating a quadratic cost function associated with a state space model of an animal possessing two essential state variables, hunger and thirst. They formulated the model in control theory and used Pontygarin's maximisation principle[2] to obtain the optimal behaviour. The formulation was

$$\dot{h} = -r_h u_h,\ \dot{t} = -r_t u_t;\ 0 \leq \tfrac{u_h}{k_h} + \tfrac{u_t}{k_t} \leq 1$$

where h represents food deficit or hunger, t represents water deficit or thirst, the r_h and r_t are the respective rates that the agent can gain resources by performing the resource satisfying behaviours, the u_h and u_t are the control variables that the agent has to set to perform the behaviours appropriate to reducing their respective deficits and the k_h and k_t are constraints upon u_h and u_t respectively which represent the maximum rate at which the behaviour can be performed. The r_h, r_t, k_h and k_t provide the model of the environment where r_h and r_t, called the availability, can be considered to be the density of the resource in the environment and k_h and k_t, called the accessibility, represent the ease with which the agent can obtain the resource through its behaviour (Tovish, 1982; Spier and McFarland, 1996). As can be seen from the equation above, u_h and u_t are variable elements and the task of the decision-making algorithm, in this case, is to set their values for every moment in time. With such a model the cost function was hypothesised to be[3]

[2] The ideas behind Pontygarin's maximisation principle are very similar to those used in dynamical programming, however, the mathematics is somewhat more tractable. Although, generally, both methods require numerical solutions. See Dixit (1976) for a clear development.

[3] The actual cost function used in Sibly and McFarland (1976) was $C = h^2 + t^2 + u_h^2 + u_t^2$, which nicely predicts the negative exponential satiation observed in operant experiments (McCleery, 1977). However, as noted in that paper, when h and t are much larger than k_h and k_t then the approximation $C = h^2 + t^2$ holds except when the agent is near satiation.

$$C = h^2 + t^2$$

which can be justified using geometric common sense by noting that as a deficit gets larger the cost of possessing that deficit increases at greater than a linear rate (however Houston and McFarland (1980) provide a more rigorous mathematical justification), although the choice of a quadratic function over any other convex function was made purely for mathematical simplicity (McCleery, 1977). This cost function has the appropriate property that the cost increases more rapidly the further away from the homeostatic equilibrium point (in this case, $h = t = 0$) the animal's essential state variables lie. Sibly and McFarland (1976) calculated the solution to this problem using Pontryagin's maximisation method and found it could be summarised as:

$$\begin{array}{c} \text{if } h.r_h.k_h > t.r_t.k_t \text{ then eat,} \\ \text{if } h.r_h.k_t < t.r_t.k_t \text{ then drink;} \end{array}$$

this solution combines the agent's state with the parameters that describe the environment. If the cost function was not quadratic but had the same accelerating and separable characteristics then the rule would build products in the form of $f(h).r_h.k_h$ where $f(h) = \frac{\partial C}{\partial h}$. As such, the structure of the rule would not change since the function would just act to scale the state variable.

At this point it is worth commenting on an assumption made when modelling the environment in this way. Essentially the environment is modelled by a ratio, $(r_h.k_h)/(r_t.k_t)$, that describes the relative rate of return that the agent would obtain from performing the behaviours. In stating this we have implicitly made the assumption that the environment is sufficiently homogeneous that, at the very minimum, the agent can achieve such a ratio before it completely depletes any essential state variable.

2.3 The Cue-Deficit model

Analysis such as in **2.2** were used to provide the intuition for a framework to explain behaviour sequencing expounded in McFarland and Sibly (1975) and based upon the idea of motivational isoclines. In this model, the deficits of the state variables combine with stimuli from the environment. Here we consider stimuli to be a cue to resources that will have consequences to the agent's state variables. Figure 2 shows a hypothetical set of lines of equal feeding tendency, or motivational isoclines (McFarland and Sibly, 1975). The decision to perform any behaviour is made by calculating the tendencies to perform all the various behaviours that the agent may exhibit and choosing the behaviour that possesses the highest tendency. It should be noted that there is no *a priori* way of determining the combination function for the deficits and cues to obtain the motivational strength - meaning that there are an infinity of ways in

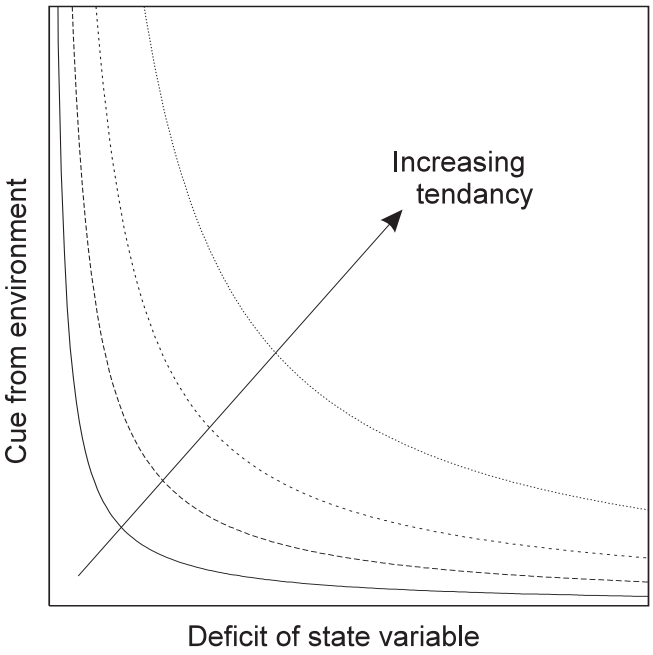

Figure 2: An example of four hypothetical motivational isoclines in cue-deficit space, any point sitting on one of the lines has an equal motivational tendency to any other on the same line. See text for details.

which the deficits and cues could be combined and no one method has reason to be better than any other. However, if we consider that the cues an agent receives from the environment about various resources allow it to obtain the potential rates of gain or loss to its state variables then we can associate the cue with, say, the $r_h.k_h$ in the quadratic cost function model.

This is attractive because under the assumption that the cue predicts a rate of gain of resource then we can hypothesise that a valid combination rule for the deficits and the cues is to multiply them together. Empirical evidence that this occurs in animals has been discussed at length in Baerends, et al. (1955), Sibly (1975) and Houston and McFarland (1976). In addition the quadratic cost function model developed in **2.2** predicts that such a multiplicative combination rule, when applied to the mechanistic definitions of the deficit and cue, should generate optimal behaviour sequencing. Of course, for this prediction to hold both the environment and the control architecture must satisfy the assumptions of the cost function model. What this means is that the multiplicative cue-deficit rule is isomorphic to the optimal solution generated in **2.2** *within* the mathematical environment. The following section investigates the performance of these mathematically isomorphic approaches when they are investigated in a spatial simulation environment. If the addition of the spatial component highlights differences between the performance of the two

rules then this may inform us about differences between planning methods (illustrated by cost function model) and motivation based reactive methods (illustrated by the cue-deficit model).

3 Our Approach

The discussion above has outlined the framework in which our work sits. The experiments described below were set out to test the control architectures developed above in a simulation environment which has a spatial component. By adding the spatial component the simple mathematical model of the environment in **2.2** no longer captures enough features of the environment to fully determine the behaviour of the agent, which has to decide from moment to moment not only what it will consume but where it will go. As such, it becomes an empirical question whether or not the inferences drawn from such a model are still valid or if the addition of a spatial component will mean the model will no longer perform as previously discussed.

A two dimensional arena in which an agent must consume resources to survive has also been investigated previously in the work of Simon (1956) and Toda (1982). Both authors have attempted to apply, to a lesser and greater extent, some mathematical formulations as their tools of investigation. Simon argued that within such an environment the main consideration for an agents control system was that it should be able to 'satisfice' its needs and was not required to make complex calculations for its decisions. Toda, on the other hand, supplied his agents with an ever more sophisticated evaluation function as a means to resolve what to do next. In some sense, the cue×deficit models are more in line with Simon's ideas whereas the cost function models follow Toda's.

In addition, we are interested in applying these ideas to robot control systems (see McFarland and Spier, 1996) so such a preliminary investigation is a halfway house to allow us to gain experience with the framework.

3.1 Experimental Model

The simulation environment consisted of a continuous toroidal surface of 10,000×10,000 units within which resided the agent and two types of resource. The resources were of 60 units radius and randomly distributed on the surface. The choice of a toroidal environment was made to avoid the requirement of designing sophisticated, and possibly confounding, navigation algorithms when the investigation was interested in the decisions the control algorithms made. In addition a walled arena would cause the resource distribution to be even less like the theoretical model. The agent was manifested as a circle of radius 60 units and possessed sensors that could identify the egocentric bearing and distance of the nearest resource of each type within the maximum range of the sensors[4] (figures 3 and 4).

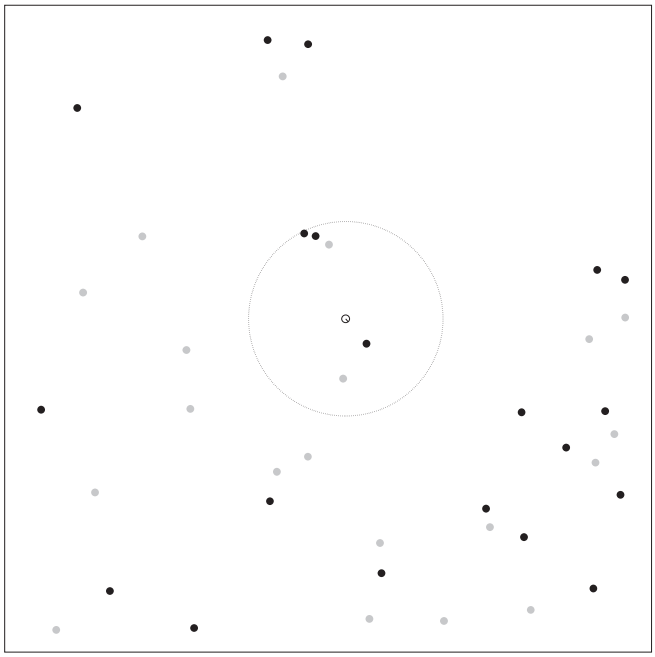

Figure 3: An example of the unclumped continuous 10000×10000 simulation environment in which the two types of resources are uniformly distributed and the agent is in the middle encircled by its maximum sensory radius of 1500.

The agent contained two internal state variables (A and B). If through its actions it permitted either to fall to zero then it would 'die'. These variables could be considered to be food and water in the case of an animal or, say, fuel and clean plates in the case of a robot. The essential idea encapsulated by this dichotomy is that the agent must maintain a homeostasis of multiple distinct state variables within itself. The state variables had a maximum value after which the agent would be effectively satiated and stop consuming that particular resource.

The agent possessed no memory, so should the agent move in a direction that takes it away from, say, a resource A towards another resource A then, when the agent gets closer to this second resource it will forget the first resource and only be able to utilise the distance and

[4] Preliminary work with this model used an agent that possessed six range sensors distributed regularly around the body. Besides the extra computational time required to implement the sensors in this way the behaviour of the agent, mediated through exactly the same control algorithms, was essentially unaffected. In addition, we are very sensitive to the possible flaws that perfect information can introduce into our model. However, one of the attractive features of some of the mechanisms we are using is that there is good reason to consider them to be noise tolerant.

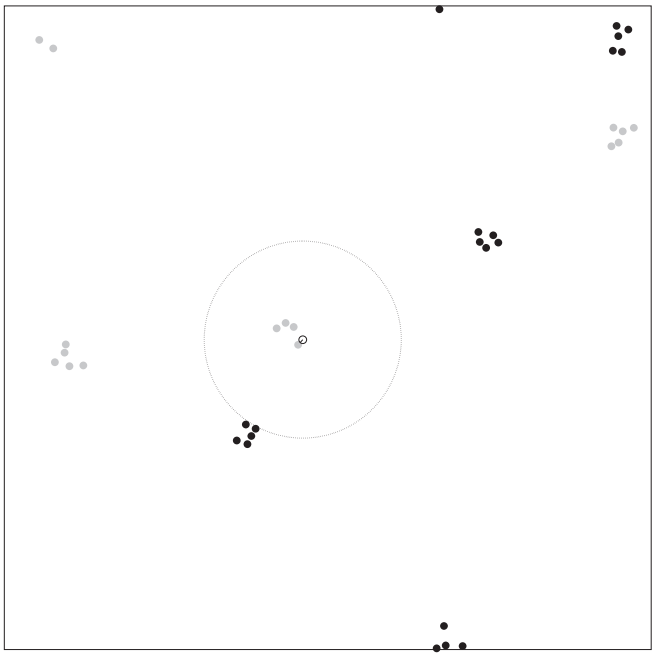

Figure 4: An example of the clumped continuous 10000×10000 simulation environment in which the two types of resources are distributed in clumps of five and the agent is in the middle encircled by its maximum sensory radius of 1500.

bearing of the second resource in its control algorithms.

While the agent was active both its state variables were diminished at a fixed ratio. When the agent was adjacent to either of the two types of resource in the environment then it could (but need not) consume them to increase the allied state variable at the cost of a smaller loss to the other internal state variable, this representing the time and/or effort used in the handling of the resource. If a resource was 'consumed' then it was randomly replaced (after a short random delay) in the environment using the distribution pattern chosen for the simulation. Three types of distributions were tested and are fully described in **4**.

The task of the agent was to survive. For measurement purposes the agent's opportunity to exist in the environment was curtailed at 50,000 cycles where a cycle consisted of either moving a standard distance in any chosen direction or, should the agent be adjacent to a resource, consuming the adjacent resource of its choice. The parameters associated with internal state expenditure per cycle and the gain per resource lump consumed were initially set so as to permit the agent a reasonable chance of surviving.

Although the mathematical model in **2.2** used only a few well specified parameters it is the nature of an environment simulator that a great many uninteresting parameters will be needed. The main criterion that specified the parameter space was that the agent should be approximately able to collect just enough resources to counter-balance the reduction in its state variables from the process of resource collection. For completeness, the parameters used within the simulation (with typical values in brackets) were: world size (10000×10000), state A loss per cycle (0.0045), state B loss per cycle (0.0045), initial internal state A (10), initial internal state B (10), maximum distance the sensors can resolve (1500), maximum value of internal state A (40), maximum value of internal state B (40), increase in internal state A after consumption of A (1), increase in internal state B after consumption of B (1), decrease in internal state B after consumption of A (0.2), decrease in internal state B after consumption of A (0.2), number of A items in the environment, number of B items in the environment, distance the agent moves in one cycle (10), A item radius (60), B item radius (60), agent radius (60).

Strategies tested within this environment were compared by slowly making the environment harder to exist in. This was performed in two ways; the strategies were titrated either by varying the loss per cycle the agent endured to its essential state variable per cycle or changing the maximum range that the agent's sensors could resolve.

3.2 Decision Strategies

By using a continuous space where the agent can effectively reside anywhere and utilising a movement step that is much smaller than the agent's own radius we avoid the traditional gridworld type of environment like that in Wilson (1985). Gridworld environments allow the control algorithms to use simplified perceptual systems that do not scale and treat time and space as discrete variables (Cliff and Bullock, 1993). Although, in this simulation we make the simplifying assumption that perception is equivalent to recognition, there are still an infinity of possible routes that an agent can adopt within the continuous environment so an exhaustive search over all such routes is an impossibility. As such, continuous space seems to provide the basis for a better model in which to explore algorithms that might be applicable to mobile robots. For this investigation the following strategies have been chosen, based upon the analysis in **2**. The criteria for the choice of the strategies was made by choosing the simplest variants of strategies based upon using the quadratic cost function and the cue×deficit rule. The rational for this was that although it is possible to carefully adjust any algorithm to perform well within the simulation environment, such an incremental approach would not provide any insights into the benefits of using the two approaches. Of course, this introduces a possible weakness of balance in this study however we feel that the strategies chosen offer a comparative range

of complexity where the question of scalability of both approaches remains open.

It should be noted, none of the strategies possessed parameters that were specifically tuned or refined for the experiments. This normal requirement was circumvented by the simplifying symmetrical nature of the environment and agent, in that, both the resource distributions and the agent's loss per cycle to its internal state variables were equal. This means that parameters typically necessary to balance between different systems were unitary.

3.2.1 Consume nearest

The first strategy is very simple, the agent moves towards the resource that is closest to it and then consumes it; should no resource be within its sensitivities, it then searches[5]. This strategy was used as a control against all the others.

3.2.2 Cue×Deficit

The second strategy combines the internal state and the external sensor information. For each of the essential state variables of the agent the motivational strength to satisfy that state is calculated by first multiplying the proximity sensor by the resource pay off to construct a cue linked to expected gain and then multiplying this complex by the current deficit each scaled between 0 and 1. The cue strength varied inversely with distance from a resource and the deficit was calculated by taking the current state away from the maximum value of the state. Competition between the two motivational tendencies decides which resource the agent pursues at any moment in time.

In intuitive notation the rule followed would be:

Go for A if
(Cue_A × A_Item_Payoff ×Deficit_A)
>
(Cue_B × B_Item_Payoff×Deficit_B)
otherwise go for B

where A_Item_Payoff and B_Item_Payoff are the respective increases in the A and B internal states of the agent should the agent consume that particular resource.

3.2.3 Cost function

The third strategy is based upon the very simple quadratic cost function[6] discussed above,

$$C = \text{A\_Deficit}^2 + \text{B\_Deficit}^2.$$

The decision to travel to either resource is formed by choosing the behaviour associated with whichever squared deficit is highest, that being the consumption of the resource that would instantaneously reduce the cost the most. If the resource cannot be seen then the agent searches.

3.2.4 One step planning cost function

A second variant of the cost function strategy is to calculate the cost function one step into the future. For example, the cost function for consuming A would be

$$\begin{aligned}\text{Cost\_A} := (&\text{A\_Deficit} - \text{A\_Item\_Payoff} + (\text{A\_Loss\_Per\_Cycle} \times \text{A\_Distance}))^2 \\ &+ (\text{B\_Deficit} - \text{B\_Loss\_From\_Eating\_A} + (\text{B\_Loss\_Per\_Cycle} \times \text{A\_Distance}))^2\end{aligned}$$

where the distance measurement is scaled to be in units of the distance the agent travels in a cycle. This means that the agent will discount the benefit of consuming the resource by the cost to its state variables of travelling to the resource to consume it. If the agent's sensors do not have a particular resource type within range then the appropriate Item_Payoff variable was set to zero. A similar cost calculation is made for consuming a B type resource and the agent sets its direction towards the resource associated with the lower overall cost. In the extreme case when one of the deficits was very close to zero the algorithm may choose to consume a resource that is out of sight, in which case the agent searches.

3.3 Reactive one step planning cost function

This strategy is the same as the one step planning cost function; however, if the appropriate resource is not in the sensory field of the agent then it will use the consume-nearest strategy.

4 Experiments

The decision strategies described above were tested in three simulated environments of differing resource distributions. Each environment contained the same global density of resource but the distribution, quantity and 'reward' of individual resource objects was varied. This equality of global density is important for the purpose of comparison because from the theoretical work in **2.3** we

[5]The search strategy is simply moving forward with a small probability (0.01) of a +/-30 degree turn to circumvent the possibility that, given all the angles in the simulation are rational, the agent gets trapped in an infinite loop moving straight ahead around the torus.

[6]Although, for simplicity, we will refer to the utility functions that the strategies use as cost functions, to be more precise we should call them goal functions. This is because a utility function that the agent can access and use to determine its behaviour is called a goal function as opposed to a cost function which is the actual function that describes the impact upon the agent from the agent-environment interaction (McFarland and Houston, 1981). An agent can possess many alternative goal functions but there can only be one cost function.

expect all the strategies (except for the consume nearest) to perform comparably in a perfectly smooth and uniform distribution of resources. Since all the environments possess the same global distribution, differences in performance will indicate how the strategies fair when tested with an unmodeled spatial component.

The first environment contained 20 objects of each resource uniformly distributed over the entire surface of the torus (figure 3). When any resource object was consumed it was regenerated (after a random short delay) in a new location found under the same uniform distribution that was initially used.

The second environment also contained 20 resource objects of each type; however, they were allocated into uniformly distributed clumps of resource, each clump containing five resource objects of the same type randomly spread within a 250 unit radius of the clump centre (figure 4). For purposes of resource regeneration, during a run the centre point of each of the current resource type clumps was stored, and as resources were consumed they were replaced within the appropriate current resource type clump until it reached its maximum allocation at which point the current clump centre was randomly allocated to a new point on the surface of the torus. Hence, although the environment was clumped, the spatial location of the clumps moved over the period of the simulation.

The third environment was the same as the first but with 40 resource objects of each type and each resource object possessing half the quantity of resource that the objects contained in the previous two environments. Although the density of resources in this environment is the same as the previous two it is to be expected that this environment would be energetically more expensive than them. The reason for this is because the actual distance the agent must travel to shuttle between all the resources has increased so although the global density was the same the travel time of the agent would have increased.

For each parameter configuration all the strategies were run for 80 trials. The cycle lifetime of the agent was recorded for each trial as a measure of the viability of the agent in the environment. Trials were curtailed at 50000 cycles. The parameters were varied over all the essential variables impinging upon the strategies, that being: the state A loss per cycle and state B loss per cycle (both altered symmetrically from 0.004 to 0.006 in steps of 0.002 unless the agents performed particularly poorly when the starting value of 0.003 was used) and the maximum distance the sensors can resolve (which ranged from 1000 to 2000 in steps of 1250 although only the ranges 1500 and 2000 are reported on here).

4.1 Results

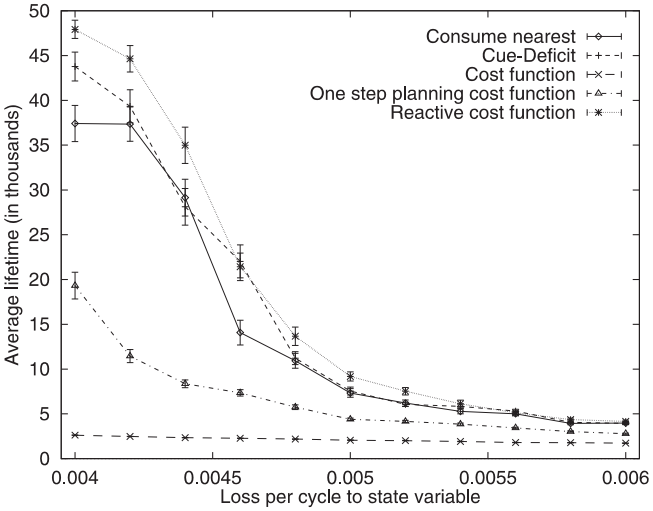

Figure 5: Results comparing the various strategies discussed in the text in the unclumped simulation environment with 20 resource objects of each type and a maximum sensory range of 1500. Each point represents a mean of 80 trials bracketed by its standard error.

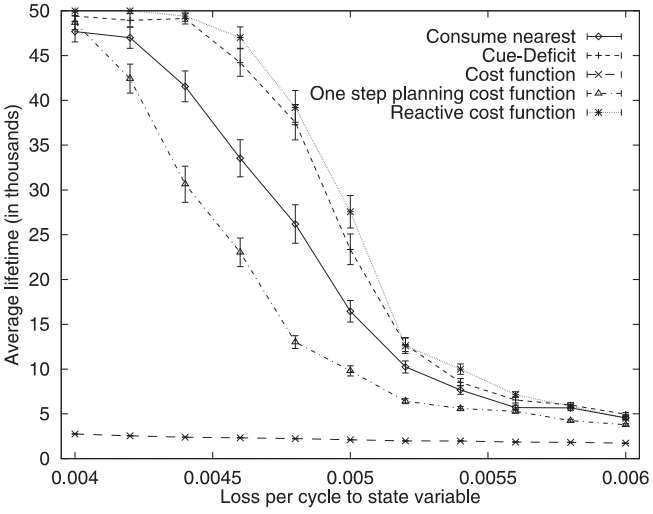

Figure 6: Results comparing the various strategies discussed in the text in the unclumped simulation environment with 20 resource objects of each type and a maximum sensory range of 2000. Each point represents a mean of 80 trials bracketed by its standard error.

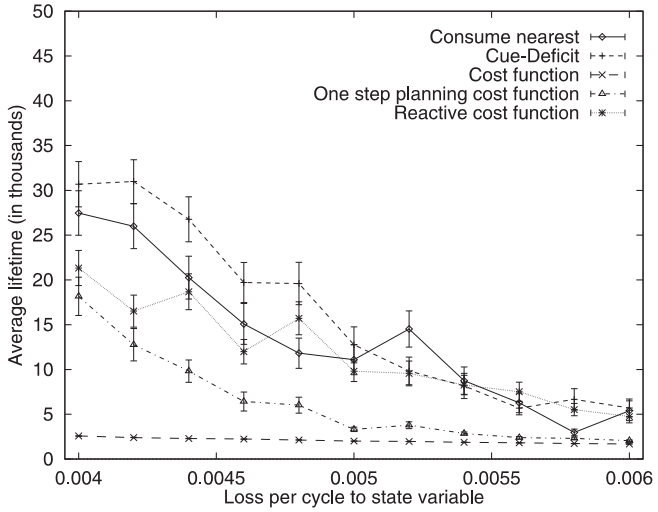

Figure 7: Results comparing the various strategies discussed in the text in the clumped simulation environment with 20 resource objects of each type in clumps of five and a maximum sensory range of 1500. Each point represents a mean of 80 trials bracketed by its standard error.

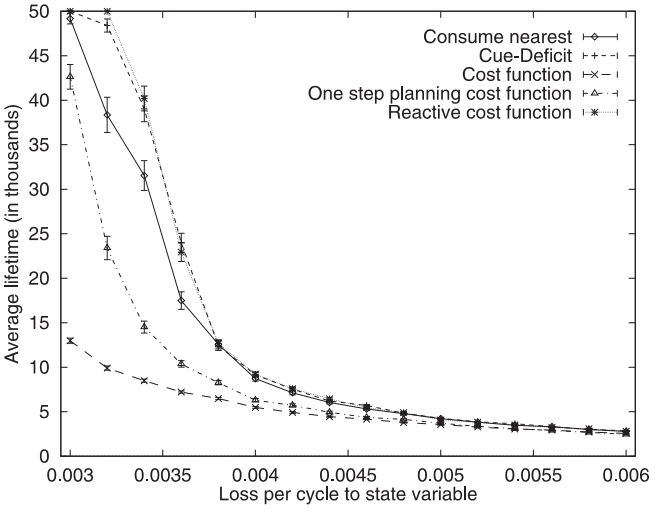

Figure 9: Results comparing the various strategies discussed in the text in the unclumped simulation environment with 40 resource objects of each type (giving half the standard resource quantity) and a maximum sensory range of 1500. Each point represents a mean of 80 trials bracketed by its standard error.

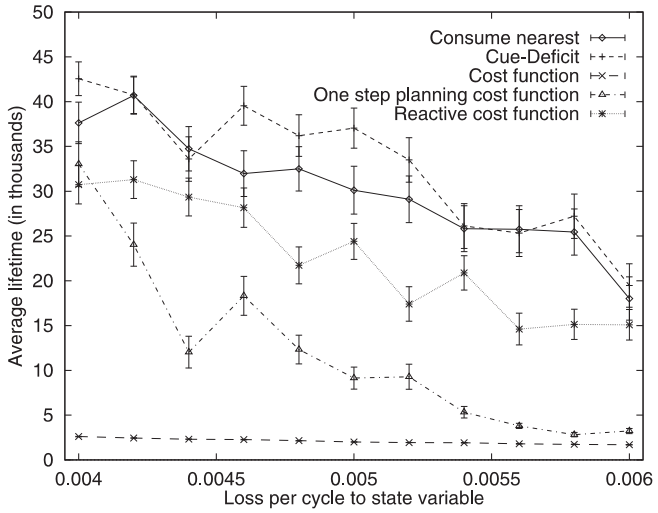

Figure 8: Results comparing the various strategies discussed in the text in the clumped simulation environment with 20 resource objects of each type in clumps of five and a maximum sensory range of 2000. Each point represents a mean of 80 trials bracketed by its standard error.

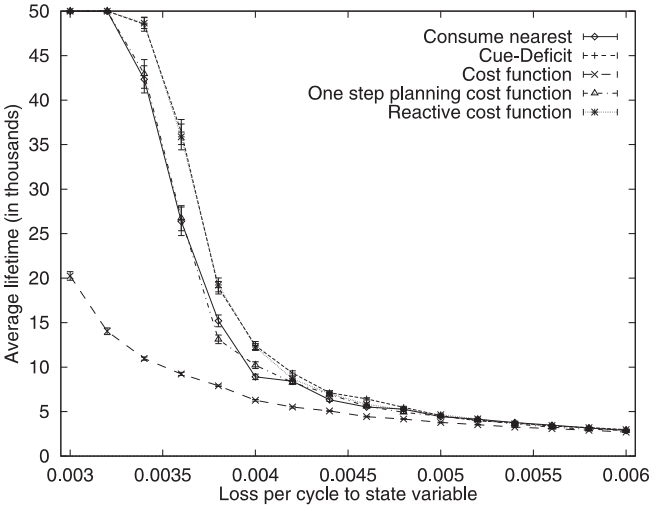

Figure 10: Results comparing the various strategies discussed in the text in the unclumped simulation environment with 40 resource objects of each type (giving half the standard resource quantity) and a maximum sensory range of 2000. Each point represents a mean of 80 trials bracketed by its standard error.

Figures 5–10 contain the results of the simulation exercise; the paragraphs below survey the differences in performance. The first thing to note is the obvious result that as the resources consumed per cycle increase, the strategies find it harder to survive. So, at very low quantities of resources consumed per cycle all the strategies will do well and any differences between strategies will not make themselves apparent. Likewise, at very high quantities of resources consumed per cycle all the strategies find it too hard to survive so, again, any performance difference is invisible. The statistical comparisons between the data took this into account by only considering trials whose mean lifetime was greater than, an arbitrarily chosen, 10000. Mood's median test (Siegel, 1988) was used to determine if two trends were statistically different by testing pairs of points recorded under the same loss per cycle parameter. Unless otherwise mentioned, all pairs on a trend, whose mean lifetime was greater than 10000, generated the same statistical result and an indicative p value of a one degree of freedom test is mentioned. The median test is a non-parametric test suitable for data, like the reported lifetime data, which is truncated.

Figures 5 and 6 show the performance of the strategies over two maximum range settings (low, 1500 and high, 2000) in an unclumped environment with 20 resource objects of each type; figures 7 and 8 show the performance of the strategies in a clumped environment with 20 resource objects of each type and figures 9 and 10 show the performance of the strategies in an unclumped environment with 40 resource objects of each type, each object possessing half the quantity of resource that the trials shown in figures 5–8 used for each object.

In the first unclumped environment (figures 5 and 6) the cue×deficit (**3.2.2**) and the reactive cost function (**3.3**) performed similarly in the high range setting (Mood, $p>0.08$) with the reactive cost function strategy performing better in the low range setting (Mood, $p<0.03$). The consume nearest strategy (**3.2.1**) performed slightly similarly to the cue×deficit in the low range setting (Mood, $p>0.25$ for two points) but worse at the higher range setting (Mood, $p<0.01$). The one step cost function strategy (**3.2.4**) performed much worse than the previous three (Mood, $p<0.0001$) and the simple cost function strategy (**3.2.3**) failed to produce any appropriate behaviour.

The clumped environment (figures 7 and 8) generated data with a high variance (discussed below) however while and consume nearest strategies performed similarly in the low range setting (Mood, $p>0.13$), in the high range setting the reactive cost function performed worse than both of the others (Mood, $p<0.02$). The one step cost function strategy performed much worse than the previous three in the high range settings (Mood, $p<0.001$) and worse than the cue×deficit and reactive methods in the low range settings (Mood, $p<0.02$). The simple cost function strategy, again, failed to produce any appropriate behaviour. Also the cost per cycle parameter range over which the strategies survived was much extended when compared to the unclumped environments.

In the unclumped environment with 40 resources of each type (figures 9 and 10) the cue×deficit and the reactive cost function performed almost exactly the same (Mood, $p>0.5$). The consume nearest strategy and the one step cost function strategy performed comparably (Mood, $p>0.5$) but with a lower average lifetime than the previous two at the high range setting (Mood, $p<0.001$). At the lower range setting the consume nearest strategy performed better than the one step cost function strategy (Mood, $p<0.001$). In both range settings the simple cost function strategy managed to perform above the minimal lifetime with its performance improving with increased range sensing.

4.2 Analysis of the results

Here we will consider each individual strategy in order of their overall performance within the simulation environments. It is important to remember that, as we argue in **2.3**, all the strategies (except the consume-nearest, which was the control) perform comparably in the constructed mathematical model of **2.2**.

4.2.1 Cost function

This strategy failed to generate any appropriate behaviour in both the unclumped, 20 resource object environment and the clumped environment. Within the unclumped, 40 resource object environment - the environment that most closely matched the mathematical models assumptions - it began to maintain a viable lifespan within the loss per cycle parameter range of the other strategies. The reason for this was its picky behaviour; it performed random search until it found the resource object that reduced its cost function the most in one go. Such insensitivity to fortuitous resource distributions would account for the low variance of its lifetimes when compared to the other strategies. The 40 resource environment increases the probability that the agent will encounter a resource item in any particular step, so reducing the search time expended between resource encounters. This permitted the cost function strategy to begin to attain comparable performance to the other strategies.

4.2.2 One step planning cost function

This strategy performed worse than the consume nearest strategy except in one configuration. Although, like the cost function strategy, it performed worse than consume

nearest when in the 20 resource object environments, within the uniform 40 resource object environment its relative performance significantly improved, equalling the consume nearest strategy at the higher range sensor value. Within the clumped environment the one step planning strategy did not exhibit a viable lifetime in the extended range of loss per cycles that the reactive strategies did. This can be accounted for by the observation that even with the one step planning component, the strategy had to dither between resource patches and so could not take advantage of the extra information it gained when it encountered a clump of resource.

4.2.3 Consume nearest

As a control, the consume nearest strategy seemed to perform admirably, separating the strategies and indicating what kind of performance a naive reactive strategy ought to be able to obtain.

4.2.4 Reactive cost function

The reactive cost function strategy performed very well in all the environments except the clumped ones when its performance was equal to or worse than the consume nearest strategy. Here we can see that the extra information the extended range sensor provided, instead of serving to improve the relative performance of the strategy, was disadvantageous to it. This is because the increased range of its sensors reduced the number of times the reactive cost function strategy could utilise its reactive component - it no longer took advantage of the opportunities to consume an entire clump of resource if, with the now more frequent occurrence, a clump of an alternative resource was within its increased site range. Instead, the higher priority cost function component of the strategy would cause the agent to shuttle between the two resource clumps - this being much less efficient behaviour.

It should be noted, here, that there was a general lower rate of change in the average lifetimes of all the strategies within the clumped environments when titrated by the loss per cycle parameter. This can be accounted for by the rational that within these environments there is a lower probability that the agent encounters any resource patch (especially when the maximum distance at which the range sensor operates is low). As a consequence, many of the agents died very early on in their lifetimes, not having had the opportunity to build up a reserve of resources within themselves. This effect would have a more serious impact on the low cost per cycle environments because the low lifetimes would have a marked impact on the mean lifetime of the otherwise high lifetimes obtained in the easy conditions. This also accounts for the higher variance in the clumped results.

4.2.5 Cue×Deficit

The cue×deficit strategy performed very well in all the environments, only being bettered by the reactive cost function in the 20 resource unclumped environment with a low sensory range.

5 Discussion

The above investigation has been an attempt to use two approaches to the problem of creating an artefact that produces sensible sequences of behaviour. A functional approach contingent upon a cost function and a mechanistic one, more closely tied to the physicality (albeit virtual) of the agent. To compare the two approaches, which are mathematically equivalent in a very simple model, we tested instantiations of the algorithms in a spatial environment. The importance of considering the spatial component is that the benefits of utilising reactive strategies become apparent when the (unknown) complexities of the world can no longer be modelled accurately[7].

As far as our particular simulation is concerned, the first thing to note is the most obvious; the strategies did not perform the same within the simulation environment. It is clear that the strategies based upon cost functions did not perform as well as the alternatives. The one exception to this is the reactive one step planning cost function strategy (**3.3**) that presumably obtained most of its additional efficacy from the consume nearest strategy (**3.2.1**). The probable cause for the cost function method's lower performance was their tendency to dither between resources, making sure one state variable did not get too far out of line of the other.

This observation can account for most of the results from the various cost function variants and is clearly demonstrated in figures 11 and 12. Figure 11 shows the cue-deficit (**3.2.2**) model taking advantage of opportunities presented to it (denoted by the **o** marks at the top of the figure), as such its state variables are permitted to drift apart somewhat. The contrary is the case in figure 12 where the cost function strategy (**3.2.4**) is seen to dither about the balance of the two state variables. The periods where it ignores obvious opportunities are marked at the top of the figure.

The downfall of the cost function strategies seems to be that they require the underlying assumption of a smooth homogeneous (and, as such, non-spatial) resource distribution in the environment, originated from the mathematical model used to stimulate their design. This occurs to varying degrees in the three environments - the 40 resource object unclumped environment being closer to the homogenous model than the other two.

[7] Rather succinctly put by Rodney Brooks (1991) as, "The world is its own best model."

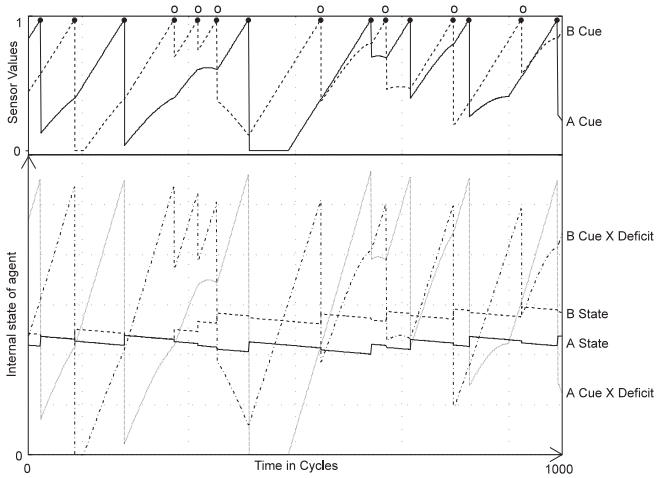

Figure 11: Unprocessed sensor and internal state data for one particular run of the cue-deficit strategy in the simulation environment showing 1000 execution cycles. The figure is divided into two halves with the particular traces labeled on the right. The top half of the figure shows the sensor values (ranging from 0 to a maximum of 1) where a consumption of a resource is marked by a bullet point. The bottom half of the figure shows the actual internal state variable for both resources as well as the cue×deficit motivational tendency. Resource consumptions denoted by an '**o**' are moments where the non-dominant resource was consumed. For this particular time-slice, state B was non-dominant for all opportunistic acts.

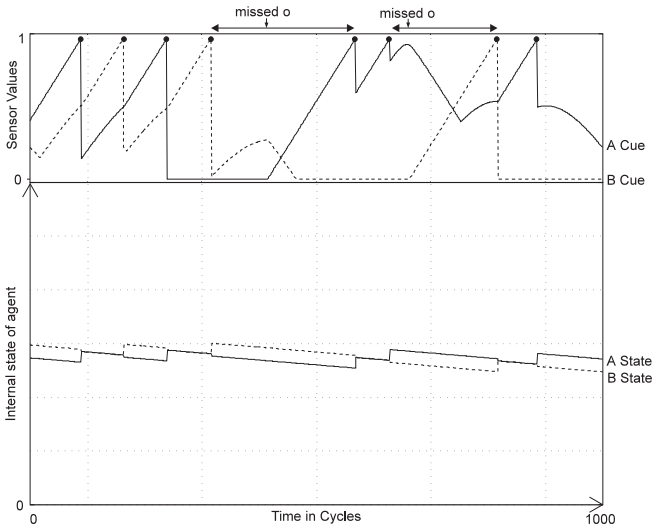

Figure 12: Unprocessed sensor and internal state data for one particular run of the cost function strategy in the simulation environment showing 1000 execution cycles. The figure is divided into two halves with the particular traces labeled on the right. The top half of the figure shows the sensor values (ranging from 0 to 1) where a consumption of a resource is marked by a bullet point. The bottom half of the figure shows the actual internal state variable for both resources. Bars labeled '**missed o**' denote periods where the agent ignores close by resource objects. The dithering behaviour of the strategy can be seen clearly.

Also, an increase in the maximum sensory range is equivalent to increasing the amount of information available to the agent and as such, somewhat improving the 'perfect' nature of the imperfect environment. For both of these reasons it is not possible to claim that the cost function which we used to model the agent can be used as a reference function with which the agent can generate optimal behaviour in the environments tested. By reference function we mean a function explicitly incorporated into the agents control system that the agent uses to guide its behaviour. It is worth noting that the reference function fails to generate anything like close to optimal behaviour despite the fact that the gross features of the environments satisfy the assumptions of the model to which the reference function matches the cost function.

This raises the question of whether it is possible to design an adequate cost function for any environment that can be placed inside an agent so it can behave appropriately. Such a cost function must take into account all the contingencies that the world places upon the agent, an infinity of situations with no reason to be continuous, and the mathematical model necessary to create such a cost function must do this also. It is not just the level of mathematical sophistication required to solve the model that is the problem, but the ability of mathematics to even describe the agent - environment interaction within a spatial dynamic time-varying world.

That is not to say that mathematics nor cost functions can be of no help to us at all. This paper demonstrates that simple models can give inspiration and justification for other mechanisms, it is just that as the worlds within which our agents interact become more complex, the capacity for rigorous mathematics to provide specific mechanisms wains. In addition work using dynamical systems theory (e.g. Beer, 1996) can provide an attractive descriptive framework which we can utilise to analyse the performance of agents but this is quite a different application of mathematics from the desire to use it to design the innards of an agents control system. Likewise, much good work has been carried out using methods associated with genetic algorithms, utilising a cost function. In these cases, the cost function acts in a shaping role to a mechanism that has no explicit access to it much as the cue×deficit strategy was related its forming cost function (e.g. Cliff et al., 1993).

As a second theme, it is worthy of note that the cue×deficit strategy consistently performed very well without so much performance deficit in the clumped environment compared to the unclumped one. Whether or not we should take reassurance in the fact that it has good reason to be an optimal strategy in a mathematical model is a different matter however, the advantage of the motivational tendency approach in the cue×deficit strategy seems to be the ability to trade-off between different levels of danger to the agent due to high deficits on its state variables yet take advantage of opportunity in situations where the risk is not so high.

6 Conclusion

From our experimental results we conclude that the cue×deficit decision strategy, an instantiation of a motivational model of behaviour, performs better than the four other strategies tested in our simulated environment. The decision strategies reflect a variety of types that can be found in the literature but do not exhaust the possibilities. We do not therefore, conclude that the cue×deficit strategy is the best possible strategy.

Our simulation environment was careful designed to take account of the animat's state, its limited perception of the outside world and the changeable nature of that world. No simulation can capture everything that is, or might be, important in the environment. While we claim that the cue×deficit strategy is the best strategy in our test environment, we recognise that some other strategy might do better in a very different environment.

Our conclusions are therefore tentative. Nevertheless, our findings are interesting because they show one strategy, derived from a tradition that eschews internal world models, does much better than two other strategies, derived from traditions that prefer to rely upon world models, mathematical or not. Moreover, the cue×deficit strategy has some support from empirical and theoretical ethological studies, it makes functional sense and it works well in mobile robots. Finally, further development along similar lines has led to ideas about new strategies that appear to exhibit even more sophisticated reactive behaviour (Spier and McFarland, 1996).

7 Acknowledgements

The authors would like to thank Seth Bullock, Herbert Roitblat for their constructive comments on earlier drafts of this paper.

References

Agre, P. and Chapman, D. (1987). Pengi : An implementation of a theory of activity. In (Eds.), *Proceedings of the Sixth National Conference on Artificial Intelligence, AAAI-87*. Morgan Kaufmann.

Baerends, G.P., Brouwer, R. and Waterbolk, H.T. (1955). Ethological Studies on Lebustes reticulatus. *Behaviour*, **8**, 249–334.

Beer, R. (1996). Towards the Evolution of Dynamical Neural Networks for Minimally Cognitive Behavior. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of*

Adaptive Behavior (ed. Maes, P., Mataric, M., Meyer, J.-A., Pollack, J. and Wilson, S. W.). Cambridge, MA: MIT Press/Bradford Books.

Bellman, R.E. (1957). *Dynamic Programming.* Princeton,NJ, Princeton University Press.

Blumberg, B. (1994). Action Selection in Hamsterdam : Lessons from Ethology. In Cliff, D., Husbands, P., Meyer, J.-A. and Wilson, S. (Eds.), *From Animals to Animats 3: Proceedings of the third International Conference on Simulation of Adaptive Behavior*, pp. 108–117. MIT Press, Cambridge, MA.

Brooks, R. (1991). Intelligence Without Reason. In *Proceedings of the 12th international conference on artificial intelligence (IJCAI-91)*, pp. 569–595.

Brooks, R.A. (1990). Elephants Don't Play Chess. *Robotics and Autonomous Systems*, **6**, 3–15.

Cabanac, M. (1992). Pleasure: the Common Currency. *Journal of Theoretical Biology*, **155**, 173–200.

Cliff, D. and Bullock, S. (1993). Adding "Foveal Vision" to Wilson's Animat. *Adaptive Behavior*, **2**, 49–72.

Cliff, D., Harvey, I. and Husbands, P. (1993). Explorations in Evolutionary Robotics. *Adaptive Behavior*, **2**, 73–110.

Dixit, A.K. (1976). *Optimization in Economics*, Oxford University Press.

Georgeff, M.P. (1987). Planning. *Annual Reviews in Computing Science*, **2**, 359–400.

Halperin, J.R.P. (1991). Machine motivation. In *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (ed. Meyer, J.-A. and Wilson, S.). Cambridge, MA: MIT Press.

Houston, A.I. and McFarland, D.J. (1976). On the Measurement of Motivational Variables. *Animal Behaviour*, **24**, 459–475.

Maes, P. (1990). Situated Agents Can Have Goals. *Robotics and Autonomous Systems*, **6**, 49–70.

Mataric, M. and Cliff, D. (1996). Challenges in Evolving Controllers for Physical Robots. *Robotics and Autonomous Systems*, **19**, 67–83.

McCleery, R.H. (1977). On Satiation Curves. *Animal Behaviour*, **25**, 1005–1015.

McFarland, D. and Nunez, A.T. (1978). Systems analysis and sexual behaviour. In *Biological Determinants of Sexual Behaviour* (ed. Hutchinson, J. B.). Chichester: John Wiley and Sons.

McFarland, D.J. and Bösser, T. (1993). *Intelligent Behaviour in Animals and Robots.* Cambridge, MA., MIT Press.

McFarland, D.J. and Houston, A.I. (1981). *Quantitative Ethology : The State Space Approach.* London, Pitman.

McFarland, D.J. and Sibly, R.M. (1975). The Behavioural Final Common Path. *Philosophical Transactions of the Royal Society (Series B)*, **270**, 265–293.

McFarland, D.J. and Spier, E. (1997). Basic Cycles, Utility and Opportunism in Self-Sufficient Mobile Robots. *Robotics and Autonomous Systems*, **in press**.

Rolls, B.J. and Rolls, E.T. (1982). *Thirst.* Cambridge, CUP.

Rosenblatt, K. and Payton, D. (1989). A Fine-Grained Alternative to the Subsumption Architecture for Mobile Robot Control. In (Eds.), *Proceedings of the IEEE/INNS International Joint Conference on Neural Networks.* IEEE.

Sibly, R. (1975). How incentive and deficit determine feeding tendency. *Animal Behaviour*, **23**, 437–446.

Sibly, R.M. and McFarland, D.J. (1974). A State Space Approach to Motivation. In *Motivational Control Systems Analysis* (ed. McFarland, D. J.), pp. 213–250. London and New York: Academic Press.

Sibly, R.M. and McFarland, D.J. (1976). On the Fitness of Behaviour Sequences. *American Naturalist*, **110**, 601–617.

Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review,* **63**, 129–138.

Spier, E. and McFarland, D. (1996). A Finer-Grained Motivational Model of Behaviour Sequencing. In *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior* (ed. Maes, P., Mataric, M., Meyer, J.-A., Pollack, J. and Wilson, S. W.), pp. 255–263. Cambridge, MA: MIT Press/Bradford Books.

Staddon, J.E.R. and Zanutto, B.S. (1996). Feeding Dynamics. In *The Functional Behaviourism of Robert C. Bolles: Learning, Motivation and Cognition.* (ed. Bouton, M. E. and Fanselow, M. S.). Washington: American Psychological Association.

Stephens, D.W. and Krebs, J.R. (1986). *Foraging Theory.* Princeton, NJ, Princeton University Press.

Toates, F. (1980). *Animal Behaviour - A Systems Approach.* New York, John Wiley & Sons.

Toates, F. (1986). *Motivational Systems.* Cambridge, Cambridge University Press.

Toda, M. (1982). *Man, Robot and Society.* MA., Martinus Nijhoff Publishing.

Tovish, A. (1982). Learning to improve the availability and accessibility of resources. In *Functional Ontogeny*

(ed. McFarland, D.). London: Pitman books.

Tyrrell, T. (1993). The Use of Hierarchies for Action Selection. In *From Animals to Animats 2: Proceedings of the Second International Conference on the Simulation of Adaptive Behavior* (ed. Meyer, J.-A., Roitblat, H. L. and Wilson, S.). Hawaii: MIT Press.

Wilson, S.W. (1985). Knowledge Growth in an Artificial Animal. In Grefenstette, J. (Eds.), *Proceedings of the first international conference on Genetic Algorithms and their Applications*, pp. 16–23. Lawrence Erlbaum Associates.

Wilson, S.W. (1991). The Animat Path to AI. In *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior* (ed. Meyer, J.-A. and Wilson, S.). Cambridge, MA: MIT Press.