Reinforcement learning

A comparison of agents in discrete environments with large discrete state spaces using reinforcement learning.

Make sure the title and text of each section match

Look over the size of all the paragraphs in the world)

Be consistent with reinforcement learning (not Reinforcement Learning, RL)

Be consistent in introductions using references to subsections

Check see section vs section, and same w figure

Tamarisk -¿ tamarix

Joakim Persson
Johan Andersson
Emil Kristiansson
Adam Sandberg Eriksson
Daniel Toom
Joppe Widstam

Abstract

Contents

Todo list

Chapter 1

Introduction

Reinforcement learning, a subfield of artificial intelligence, is the study of algorithms that learn how to choose the best actions depending on the situation, that is, the state of the environment. The best action or sequence of actions is what leads to the best results, that is to say, the greatest rewards. In a reinforcement learning problem the algorithm is not told which actions maximize the reward, instead it has to interact with the environment to learn when and where to take a certain action. When the agent has learnt about each situation or state of the environment, it will arrive at an optimal sequence of actions. (Barto and Sutton 1998).

Reinforcement learning can be exemplified by how a newborn animal learns how to stand up. There is no tutor to teach it, so instead it tries various combinations of movements, while remembering which of them lead to success and which of them lead to failure. Probably, at first, most of its attempted movement patterns will lead to it falling down - a negative result, making the animal less likely to try those patterns again. After a while, the animal finds some combination of muscle contractions that enables it to stand - a positive result, which would make it more likely to perform those actions again.

In the real world, what actions lead to success almost always depends on the circumstances in which they are taken. For instance, a person might be rewarded if they sang beautifully at a concert, but doing the same at the library would probably get them thrown out. Furthermore, for many real world domains it is common that the state space (the set of possible states of the environments in which actions can be taken) for the reinforcement learning problem becomes very large (Guestrin et al. 2003). To continue our example, there are probably countless of places and situation where it is possible to sing. In this case, how does one abstract and find the qualities of the environment that are important for deciding on the optimal action?

For an algorithm used by a computer, there are two problems connected to large state spaces. First, it is hard to store representations of the states of the environment in main memory (Szepesvári 2010). Second, it takes too long to repeatedly visit all the states to find the best action to take in all of them

(Dietterich, Taleghan, and Crowley 2013).

1.1 Purpose and problem statement

This thesis aims to further the knowledge of reinforcement learning or more specifically, algorithms applied to reinforcement learning problems with large state spaces. Reinforcement learning algorithms struggle with environments consisting of large state spaces due to practical limitations in memory usage and difficulties in estimating the value of actions and states within reasonable processing time. Therefore, we aim to evaluate techniques to circumvent or reduce the impact of these limitations.

1.2 Limitations

Due to time constraints only two algorithms are tested, which apply different techniques. Furthermore, the evaluation of the selected algorithms is done in only one environment covering a certain problem with a large state space.

1.3 Simulation environment

The environment used is Invasive Species from the Reinforcement Learning Competition, described in section 4.2. Depending on the parameters, the problem can have both a large state space and/or a large action space. This makes Invasive Species a good choice with the given problem statement.

1.4 Agents to be evaluated

Two algorithms that are suitable to the purpose and problem statement, since they apply different techniques for working with large state spaces, are Explicit Explore or Exploit in Dynamic Bayesian Networks (DBN-E^3) and Model Based Interval Estimation (MBIE). They are described in detail in chapter 3.

Chapter 2

Technical background

This chapter is for readers new to the reinforcement learning area. A short overview is first given of artificial intelligence and how reinforcement learning fits within the field of artificial intelligence, the chapter then continues with the main topics and concepts necessary to understand the rest of the report.

2.1 Artificial intelligence

The term artificial intelligence was coined by John McCarthy who described it as "the science and engineering of making intelligent machines" (McCarthy 2007). More specifically, it addresses creating intelligent computer machines or software that can achieve specified goals computationally. These goals can comprise anything, e.g. writing poetry, playing complex games such as chess or diagnosing diseases. Different branches of artificial intelligence include planning, reasoning, pattern recognition and learning from experience.

2.2 Reinforcement learning

Reinforcement learning is an approach to the learning from experience problem in artificial intelligence. A reinforcement learning algorithm uses past experiences and domain knowledge to make intelligent decisions in the future (Barto and Sutton 1998).

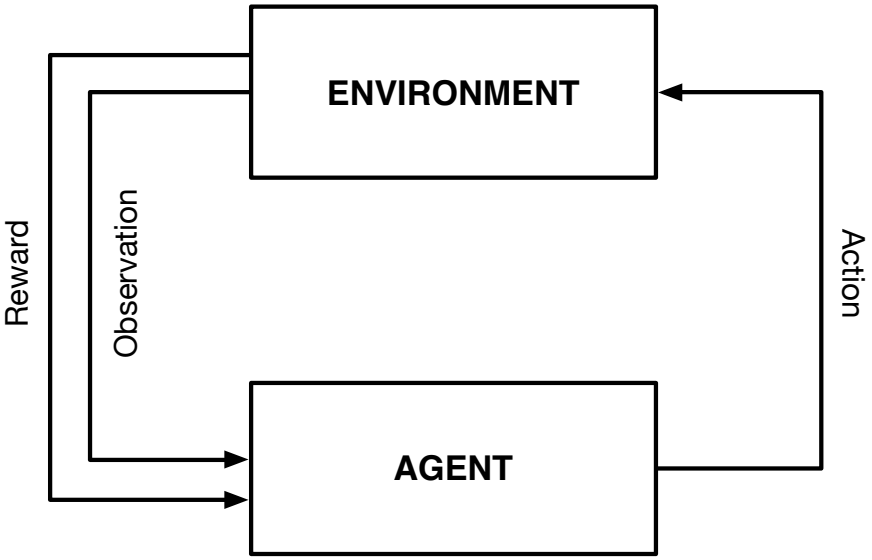


Figure 2.1: The reinforcement learning process

The reinforcement learning problem is modeled as a sequential decision problem, see figure 2.1 for a graphical representation of the process. A learning agent performs an action and receives a reward according to some measure of how desirable the results of the action are. After the action is taken, the state of the environment changes and the agent then receives a new observation and the process repeats. The goal of the reinforcement learning agent is to maximize the reward received over a certain time period, finding a balance between immediate and future rewards (Barto and Sutton 1998).

There are several ways of further dividing reinforcement learning problems into other subcategories. Some examples are the dichotomies: episodic/non-episodic problems, problems with continuous/discrete state spaces, problems with continuous/discrete action spaces, problems with one/multiple concurrent agents etc. In the text below, the two first of these dichotomies, which are relevant to this thesis, are discussed.

2.2.1 Episodic and non-episodic problems

One can categorize reinforcement learning problems based on whether or not the time steps are divided into episodes. When the time steps for the agent-environment interaction are divided into subsequences in this manner, the problem is called episodic. This is common in for example games, which end when they are won or lost. If the problem is not episodic, it is called non-episodic, which means, consequently, that the interaction is not divided into subsequences. Instead, the interaction between the actor and environment goes on continually

without end (Barto and Sutton 1998). Two examples of non-episodic applications are controlling a robot arm (as well as other robotics problems) and maneuvering a helicopter (Ng et al. 2006).

2.2.2 Continuous and discrete problems

An alternative way to categorize reinforcement learning problems is based on whether their state spaces are continuous or discrete. An important difference between the two kinds of problems is how an agent can treat similar states in the model. In a continuous problem it is a lot easier to group together states located around the same "position" due to the nature of a continuous problem, where the states usually more or less meld together. For example, if the reinforcement learning problem is to control a robot arm whose position is the state of the environment, the different states located around the same general area are not very different. This means that they might be treated in the same way or a similar way by an agent. On the other hand, consider the discrete problem of a game of connect four, wherein the placement of a coin into one of two adjacent slots could dramatically change the evaluation of the state and the outcome of the game (Barto and Sutton 1998).

2.3 Markov decision process

Within reinforcement learning, the concept of Markov Decision Processes is central. A Markov decision process (MDP) is a way to model an environment where state changes in it are dependent on both random chance and the actions of an agent. An MDP is defined by the quadruple $(S, A, P(\cdot,\cdot,\cdot), R(\cdot,\cdot))$ (Altman 2002):

S
: A set of states representing the environment.

A
: A set of actions that can be taken.

$P\colon S \times A \times S \to \mathbb{R}$
: A probability distribution over the transitions in the environment. This function describes the probability of ending up in a certain target state when a certain action is taken from a certain origin state.

$R\colon S \times A \to \mathbb{R}$
: A function for the reward associated with a state transition. In some formulation of MDPs the reward function only depends on the state.

MDPs are similar to Markov chains, but there are two differences. First, there is a concept of rewards in MDPs, which is absent in Markov chains. Second, in a Markov chain, the only thing that affects the probabilities of transitioning to other states is the current state, whereas in an MPD both the current

state and the action taken in that state are needed to know the probability distribution connected with the next state (Altman 2002).

2.3.1 Markov property

A defining characteristic of an MDP is the Markov property - that, given the current state of the environment, one cannot gain any more information about its future behavior by also considering the previous actions and states it has been in. This can be compared to the state of a chess game, where the positions of the pieces at any time completely summarizes everything relevant about what has happened previously in the game. That is, no more information about previous moves or states of the board is needed to decide how to play or predict the future outcome of the game (disregarding psychological factors). A chess MDP that uses a chess board as its state representation could thus be an example of an MDP with the Markov property (Altman 2002).

2.3.2 Sparse MDPs

If there are only a few states s' for which $P(s, a, s') > 0$, the MDP is called a sparse MDP. That is to say, when an agent performs a certain action, a, in a certain state, s, the environment can only end up in a small fraction out of the total number of states. If an MDP is sparse, there are several possible optimizations that can be performed (Dietterich, Taleghan, and Crowley 2013). The MBIE algorithm (section 3.1) can be extended to utilize such optimizations, one of which is described in section 3.1.3.

2.3.3 Representations

Two ways to separate the representations of an MDP is extensional or factored representation. Depending on the problem domain it can be advantageous from the computational view to use factored representation (Boutilier, Dean, and Hanks 1999).

Extensional representation The most straightforward way to model an MDP is called extensional representation, where the set of states and actions are enumerated directly. It is also commonly refered to as an explicit representation and closely mirrors the definition we have used so far in the report when discussing the abstract view of an MDP (Boutilier, Dean, and Hanks 1999).

Factored representation A factored representation of the states of an MDP often results in an more compact way of describing the set of states. Certain properties or features of the states are used to categorize the states into different sets. Then one can treat all the members of the same set in the same manner. Which properties or features are used is chosen by the algorithm designer, to fit the environment.

When the MDP is factored, it enables a factored representation of rewards, actions and other components of the MDP as well. When using a factored action representation, an action can be taken based on specific state features instead of on the whole state. If the individual actions affect relativity few features or if the effects contain regularities then using a factored representation can result in compact representations of actions (Boutilier, Dean, and Hanks 1999).

2.4 Basic algorithms for solving MDPs

A policy π is a function from a state s to an action a that operates in the context of a Markov Decision Process, i.e. $\pi \colon S \to A$. A policy is thus a description of how to act in each state of an MDP. An arbitrary policy is denoted by π and the optimal policy (the policy with the largest expected reward in the MDP) is denoted by $*$. The rest of this section is describes some basic algorithms for solving, that is to say finding an optimal policy, in an MDP.

2.4.1 Value functions

To solve an MDP most algorithms make uses an estimation of values of states or actions. Two value functions are usually defined, the state-value function $V : S \to \mathbb{R}$ and the state-action-value function $Q : S \times A \to \mathbb{R}$. As the names implies V signifies how good a state is, while Q signifies how good an action in a state is. The state-value function, $V^\pi(s)$, returns the expected value when starting in state s and following policy π thereafter. The state-action-value function $Q^\pi(s, a)$ returns the expected value when starting in state s and taking action a and thereafter following policy π. The value functions for the optimal policy are denoted by $V^*(s)$ and $Q^*(s, a)$.

Both V^π and Q^π can be estimated from experience. This can be done by maintaining the average of the rewards that have followed each state when following the policy π. When the number of times the state has been encountered goes to infinity, the average over these histories of rewards converges to the true values of the value function (Barto and Sutton 1998).

2.4.2 Dynamic programming

Dynamic programming is a way of dividing a problem into subproblems that can be solved independently. If the result of a particular subproblem is needed again, it can be looked up from a table. In reinforcement learning, dynamic programming can be used to calculate the value functions of an MDP (Bellman 1957). Examples of dynamic programming algorithms are policy iteration and value iteration, which are discussed in sections 2.4.3 and 2.4.4. These algorithms are often the basis for more advanced algorithms, among them the ones described in chapter 3.

2.4.3 Policy iteration

Policy iteration is a method for solving an MPD that will converge to an optimal policy and the true value function in a finite number of iterations, if the process is a finite MDP. The algorithm consists of three steps: initialization, policy evaluation and policy improvement (Barto and Sutton 1998).

Initialization
Start with an arbitrary policy π and arbitrary value function V.

Policy evaluation
Compute an updated value function, V, for policy π in the MDP using the update rule (2.2).

Policy improvement
Improve the policy by making it greedy with regard to V.

Repeat evaluation and improvement until π is stable between two iterations.

Policy evaluation is carried out using the update rule (2.1) until V converges policy improvement uses the update rule (2.2)[1].

$$V_{k+1}(s) = \sum_{s'} P(s, a, s') \left[R(s, s') + \gamma V_k(s')\right] \tag{2.1}$$

$$\pi_{k+1}(s) = \arg\max_a P(s, a, s') \left[R(s, s') + \gamma V_k(s')\right] \tag{2.2}$$

2.4.4 Value iteration

Value iteration is a simplification of policy iteration where only one step of policy evaluation is performed in each iteration (Barto and Sutton 1998). Value iteration does not compute an actual policy until the value function has converged. Value iteration works as follows:

Initialization
Start with an arbitrary value function V.

Value iteration
Update V for each state using the update rule (2.3)

Repeat value iteration until V converges

Compute the policy using (2.4)

$$V_{k+1}(s) = \max_a \sum_{s'} P(s, a, s') \left[R(s, a) + \gamma V_k(s')\right] \tag{2.3}$$

$$\pi(s) = \arg\max_a \sum_{s'} P(s, a, s') \left[R(s, a) + \gamma V_k(s')\right] \tag{2.4}$$

[1] $\arg\max_a f(a)$ gives the a that maximizes $f(a)$.

2.5 Dynamic Bayesian networks

A Bayesian network is a graphical model that represents random variables and their dependencies on eachother (see figure 2.2). Each random variable corresponds to a node and a directed edge represents a dependency between the target and the source node. (Heckerman 1998). For instance in figure 2.2, the probabilty distributions for the variable z can be changed when the variables x and y are observed.

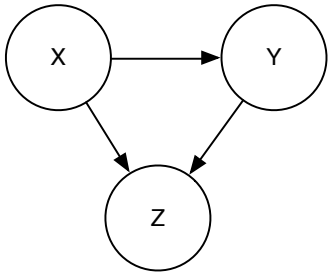


Figure 2.2: A simple Bayesian network

A dynamic Bayesian network (DBN) is a Bayesian network where the random variables are allowed to depend on prior settings of the same random variables as well as eachother, see figure 2.3.

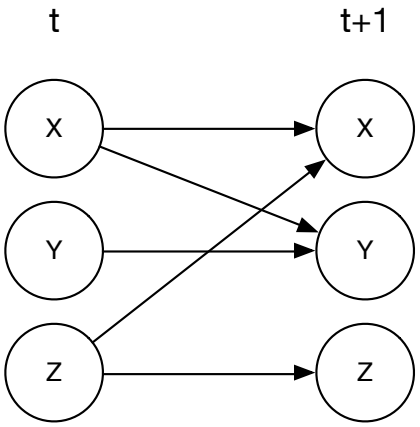


Figure 2.3: A simple dynamic Bayesian network

Transition probabilities for the states of an MDP can be represented by a set of Bayesian networks, one network for each possible action. Such a network is

comprised of a node, n_i, for each state variable, representing that state variable at time t, as well as a node, n'_i for each state variable representing the value of that state variable at time $t + 1$. If the probability distribution of a certain state variable n'_k is affected by the value of another state variable n_j if action a is taken then there is a directed edge from n_j to n'_k in the DBN corresponding to action a (Guestrin et al. 2003).

Chapter 3

Algorithms

The main topic of the thesis is to study techniques that can be used for reinforcement learning in large state spaces. Both algorithms in this chapter deal with this issue; however, they differ in the methods they apply. In the following chapter the general ideas behind the algorithms, as well as specific details, are presented.

The model-based interval estimation algorithm, described in section 3.1, utilizes clever estimations of confidence intervals for the Q-value functions to improve performance in sparse MDPs. Section 3.2 is on an algorithm that uses dynamic Bayesian networks and factored representations to improve the E^3 algorithm to more efficently deal with factored MDPs.

3.1 Model-based interval estimation

Model-based interval estimation modifies value iteration with confidence intervals. These confidence intervals allow the agent to choose between actions, based on how confident it is about its evaluation of them. In effect, the less certain the agent is about its evaluation of the states and actions, the more exploratory the actions will be. When the agent is more confident however, it will exploit what it has learnt so far about the MDP (Dietterich, Taleghan, and Crowley 2013).

3.1.1 Value iteration with confidence intervals

The confidence bounds on the Q-values in the MBIE-algorithm are calculated by making a maximally optimistic estimation of these values, given some confidence parameter. The less times a state-action pair has been visited, the more optimistic this estimation will be. This has the effect of promoting exploration of actions that have been taken few times.

When a state is first encountered by the agent, the Q-values associated with the state are initialized with the maximum achievable reward. When the ac-

tions are later performed, the state-action pairs have their Q-values gradually decreased depending on the expected value. Given time, the confidence bounds will become smaller and smaller, and the policy will converge to optimal actions with confidence specified by a confidence parameter. The bound for the confidence interval on a Q-value can be calculated by iterating the following equation (cf. section 2.4.4 about the basic value iteration algorithm) for all state-action pairs until it converges:

$$Q_{upper}(s,a) = R(s,a) + \max_{\tilde{P}(s,a) \in CI(P(s,a),\delta_1)} \gamma \sum_{s'} \tilde{P}(s'|s,a) \max_{a'} Q_{upper}(s',a') \tag{3.1}$$

To efficiently calculate the correct expected maximum Q-values in equation (3.1), a method is used that generates a distribution of the transition probabilities $\tilde{P}$ that maximizes the sum in the equation (Strehl and Littman 2008). This algorithm is referred to as Compute$\tilde{P}$ in this report.

3.1.2 Compute$\tilde{P}$, optimistic estimations of transition probabilities

The fundamental idea of the Compute$\tilde{P}$ method is that it starts with the observed transition probabilities $\hat{P}$ and then it moves probability mass from "bad" outcomes to "good" outcomes and finally returns an updated version of $\tilde{P}$. This updated $\tilde{P}$ maximizes the sum in equation (3.1), which has been proved in Strehl and Littman (2008). The state transition probability distribution is initialized according to (3.2), which corresponds to the observed probabilities.

$$\tilde{P} := \hat{P} = \frac{N(s,a,s')}{N(s,a)} \tag{3.2}$$

In (3.2), $N(s,a)$ is the number of times action a has been taken in state s and $N(s,a,s')$ is the number of times action a has been taken in state s and the agent ended up in state s'.

The procedure of moving probabilities is done by first finding the outcome state with the best V-value and observed probabilty less than 1, calling it $\overline{s}$. Analogously the outcome with the worst V-value with an observed probability of greater than 0 is found, this state is called $\underline{s}$.

The probability values $\tilde{P}(\underline{s}|s,a)$ and $\tilde{P}(\overline{s}|s,a)$ are then increased or decreased according to equations (3.3) and (3.4).

$$\tilde{P}(\underline{s}|s,a) := \tilde{P}(\underline{s}|s,a) - \xi \tag{3.3}$$

$$\tilde{P}(\overline{s}|s,a) := \tilde{P}(\overline{s}|s,a) + \xi \tag{3.4}$$

Since we need to ensure that the sum of the probabilities sum to one and that no single transition probability falls below zero or exceeds one we are only

allowed to modify the probability distribution by at most ξ, as given by equation (3.5).

$$\xi = \min\{1 - \tilde{P}(\overline{s}|s,a), \tilde{P}(\underline{s}|s,a), \Delta\omega\} \tag{3.5}$$

$$\Delta\omega = \frac{\sqrt{\frac{2|ln(2^{|S|}-2)-ln\delta|}{N(s,a)}}}{2} \tag{3.6}$$

$|S|$ denotes the total number of states and $\Delta\omega$ denotes the total probability mass to be moved. If ξ is less than $\Delta\omega$, new states $\overline{s}$ and $\underline{s}$ are found, and probabilities moved until mass equal to $\Delta\omega$ has been moved in total.

The confidence interval $CI(\hat{P}|N(s,a),\delta_1)$ (equation (3.7)) denotes a set of probability distributions where the probability is $1-\delta$ for each element in that set where the confidence interval is within distance of $\omega(N(s,a),\delta)$ of the maximum likelihood estimate for P (Dietterich, Taleghan, and Crowley 2013). This means that with probability $1-\delta$, the actual transition probabilties are between the observed probabilities and the probabilities returned by Compute$\tilde{P}$.

$$CI(\hat{P}|N(s,a),\delta_1) = \left\{\tilde{P} \,\middle|\, \|\tilde{P} - \hat{P}\|_1 \leq \omega(N(s,a),\delta)\right\} \tag{3.7}$$

Is the last sentence correct?

3.1.3 Optimizations based on Good-Turing estimations

One problem with the method described above is that probability mass can be moved to any outcome state, without any consideration taken as to whether this outcome has ever been observed. Dietterich, Taleghan, and Crowley (2013) make use of an optimization that deals with this by limiting the probability mass that can be moved to outcomes that have never been observed. The limit that is used is the approximation of the probability mass in unobserved outcomes as estimated by Good and Turing as $\hat{M}_0(s,a) = |N_1(s,a)|/N(s,a)$ (Good 1953). In this equation $N_1(s,a)$ is a set of the states that have been observed exactly once as an outcome when taking action a in state s and $N(s,a)$ is the number of times that action a has been taken in state s in total.

3.1.4 How often to perform planning

It is possible to perform planning and compute a new policy once for each action taken by the agent. However, this would be unnecessarily slow to compute. The planning comprises iterating Q-value updates to convergence and then using these converged values to update V-tables, a considerable number of computations. So instead of planning after every action taken, the algorithm only performs planning and updates the policy at some given interval.

A way to do this that we have used is to only perform an update when the number of times a specific state-action pair has been visited has doubled. For larger variants of the invasive species environment we perform planning when the number of actions taken has been multiplied by 1.5.

3.1.5 Optimizing bounds

Another optimization that can be performed is that the value $\Delta\omega$ in equation (3.5) can be tweaked to fit the environment that the agent is used with. Equation (3.5) gives bounds for which it can be proved that the method always converges to an optimal policy. In practice, however, this value can be reduced by quite a bit in order to speed up the rate at which the agent considers state-action pairs known.

A simple linearly declining function can be used instead of equation (3.5). In the so called realistic implementation of MBIE we have used $\omega = 1 - \alpha N(s, a)$. The value of the α parameter was decided through experimentation (see section 4.3).

3.2 E^3 in factored Markov Decision Processes

The second algorithm studied in this thesis is a version of the E^3 algorithm that focuses on factored problem domains by modeling them as a dynamic Bayesian network. The original E^3 algorithm is described in section 3.2.1, which gives a broad overview along with the key strategies used in the algorithm. The following section, 3.2.2, considers some ways to extend the original algorithm and make use of factored representations and planning in factored domains to improve the running time of the algorithm.

3.2.1 The E^3 algorithm

E^3 (Explicit Explore or Exploit) is an algorithm that divides the state space into two parts - known states and unknown states - in order to decide whether it is better to explore unknown states or to exploit the agent's knowledge of the known states. A state is considered to be known if the E^3 agent has visited it enough times. All other states are either unknown or have never even been visited. Unknown and unvisited states are treated in the same way. In the following sections a description is given of the three phases of operation of E^3 which are called balanced wandering, exploration and exploitation (Kearns and Singh 2002).

Balanced wandering When the agent finds itself in a state that it has not visited a large enough number of times to be considered a known state, it enters a phase called balanced wandering. When in balanced wandering, the agent always takes the action performed from this state the least number of times.

Exploration When the agent from the balanced wandering phase enters a state that is known, it performs a policy computation to find a policy that maximizes the agent's chance of ending up in an unknown state.

This exploration policy calculation is performed on an MDP which contains all known states and their experienced transition probabilites. All unknown

states are gathered in a super-state with transition probability 0 to all known states and 1 to itself. The rewards are set to 0 for known states whereas the reward for the super-state is set to the maximum possible reward. A policy based on this MDP definition will strive to perform actions that reach the super-state, i.e., an unknown state.

If the chance of ending up in the super-state is below a certain threshold, it can be proved that the agent knows enough about the MDP that it is probable that it will be able to calculate a policy that is close to optimal (Kearns and Singh 2002).

Exploitation Inversely, if the agent's chance of being able to explore is low enough, it performs a policy computation in order to find a policy that maximizes rewards from the known part of the MDP. This exploitation policy computation is performed on an MDP comprising all known states, their observed transition probabilities and their observed rewards. A super-state representing all unknown states is also added to the MDP with reward 0 and transition probability 0 to all known states and 1 to itself. This MDP definition will result in a policy that favors staying in the known MDP and finding a policy with high return.

Leaving the exploitation and exploration phases When the agent is in either the exploration or exploitation phase, there are two events that can trigger it to exit these phases. First, if the agent enters an unknown state, it goes back to the balanced wandering phase. Second, if it has stayed in the exploration or exploitation phase for T timesteps, where T is the horizon time for the MDP, it goes back to the behavior described in the "exploration" section above.

3.2.2 Factored additions to E^3

The E^3 algorithm does not exploit that the underlying Markov Decision Process may be structured in a way that allows certain optimizations. Therefore E^3 has a running time that scales polynomially with the number of states in the MDP. However, by using a factored approach for the problem, improvements can be made to the running time. By factoring the problem as a dynamic Bayesian network, the running time will scale with the number of random variables in the underlying DBN instead for the number of states (Kearns and Koller 1999).

When using a factored representation some changes to the original algorithm are required to make it compatible. One issue that has to be solved is how to perform planning with the new representation. This thesis a modified version of value iteration was used for planning and it is described later in this section. In section 7.2.2 there are other methods presented.

Dynamic Bayesian network structure Assume that the states of an MDP each are divided into several variables. For instance, the invasive species MDP described in section 4.2 constitutes such a case, where the status of each reach

can be considered a variable on its own. The number of tamarisk trees, native trees and empty slots in a certain reach at time step $t + 1$ depends not on the whole state of the environment at time t, but only on the status of adjacent reaches. Those variables on which another variable depend are called its parents.

An MDP that follows the description in the previous paragraph is described as factored. With the assumption of a factored MDP, it is possible to describe its transition probabilities as a dynamic Bayesian network, where one would have a small transition probability table for each of the reaches in the MDP, instead of a large table for the transitions for the whole states.

Planning in dynamic Bayesian networks The DBN-E^3 algorithm does not in itself define what algorithm should be used for planning when the MDP is structured as a DBN (Kearns and Koller 1999). It considers planning a black box, leaving the choice of planning algorithm to the implementers.

Value iteration can be done with a factored representation of an MDP in a fairly straightforward manner. The same equations that normal value iteration (section 2.4.4) is based on can be used when the MDP is factored too. The only difference is that in order to calculate the probability of a state transition, $p(s'|a, s)$ one has to find the product of all the partial transitions,

$$\prod_i p(s'_i|a, s_{pa(i)}) \tag{3.8}$$

where i ranges over all partial states and $s_{pa(i)}$ is the setting of the partial states that have an influence on the value of s'_i.

When an MDP has this structure, observations of partial transitions can be pooled together when the state variables are part of similar structures in the MDP. In the version of DBN-E^3 described here, all state variables that have the same number of parent variables have their observations pooled together.

3.2.3 One policy per state variable

For some MDPs it is possible to create a separate policy for each state variable individually. This is the case when there is a separate action taken for each state variable, which is true for the invasive species environment (section 4.2). In the implementation of E^3 used in this thesis, this policy creation is performed in two steps.

Planning for each state variable individually has the benefit of making the planning algorithm linear in the number of state variables, greatly reducing the time until an agent can enter the exploitation phase for large state spaces. However, there are several downsides to using this kind of approximation, some of which are discussed in section 6.1.1.

Chapter 4

Method

This chapter covers the preliminaries and preparations carried out before the execution of the experiments. It covers how the agents described in chapter 3 were implemented and it describes the environment used in the experiments, with its characteristics and test specifications. The chapter concludes with a description of the tools used for the experiment.

4.1 Algorithm implementation

The experiment utilizes RL-Glue (section 4.5) for connecting the agents and environment for the experiment. Thereby using the pre-defined specifications of RL-Glue when developing the agents, improving their re-usability.

The most complex phase of constructing the agents was the verification of their behaviour. The difficulty of this is correlated with the size of the problem domain and for a smaller problem the behaviour is easier to verify. For this environments were constructed specifically for testing the agents, an important design consideration was to make sure that the correct results were easy to either derive or were obvious from inspection. By starting with smaller problems and using an iterative approach it was possible to identify bottlenecks in our implementations and correct possible errors early. The two following sections describe the GridWorld and network simulator enviornments that were used for this purpose.

4.1.1 GridWorld

The GridWorld environment was implemented to easily be able to verify the correctness of the MBIE algorithm. It consists of a grid of twelve squares with one blocked square, one starting square, one winning square, one losing square and eight empty squares. The agent can take five actions, north, south, west, east or exit. The exit-action is only possible from the winning or losing state. When taking an action being in one state and the action is directed to another

empty state there is an 80% probability to succeed and 10% probability to fail and 10% to go sideways.

4.1.2 Network simulator

A simple computer network simulation was implemented as a simple test for the E^3 algorithm. In this environment, the agent tries to keep a network of computers up and running. All computers start in the running state, but there is a chance that they randomly stop working. If a computer is down, it has a chance to cause other computers connected to it to also fail. In each time step, the agent chooses one computer to restart, which with 100 percent probability will be in working condition in the next time step. The agent is rewarded for each computer in the running state after each time step.

4.2 Environment specification

For the experiment, the invasive species environment from the 2014 edition of the Reinforcement Learning Competition was used. The environment is a simulation of an invasive species problem, in this case a river network with invading species where the goal of the agent is to eradicate unwanted species while replanting native species.

The environment's model of the river network has parameters, such as the size of the river network and the rate at which plants spread, which can be configured in order to create different variations of the environment. The size of the river network is defined by two parameters: the number of reaches and the number of habitats per reach. A habitat is the smallest unit of land that is considered in the problem. A habitat can either be invaded by the tamarix, which is an unwanted species, empty or occupied by native species. A reach is a collection of neighboring habitats. The structure of the river network is defined in terms of which reach is connected to which (Taleghanand, Crowley, and Dietterich 2014). In figure 4.1 a model of a river network is shown.

There are four possible actions (eradicate tamarisks, plant native trees, eradicate tamarisks and plant native trees, and finally a wait-and-see action), and the agent chooses one of these actions per reach per time step. What actions are available to the agent depends on the state of each reach. It is always possible to choose the wait-and-see action, but there has to be one or more tamarisk-invaded habitats in a reach for the eradicate or eradicate and plant actions to be available and there has to be at least one empty habitat in a reach for the plant-native-trees action to be avalable (Taleghanand, Crowley, and Dietterich 2014).

4.3 Test specification

The testing of the agents required us to choose certain sets of parameters, for the environment, the two different agents and the experiment itself.

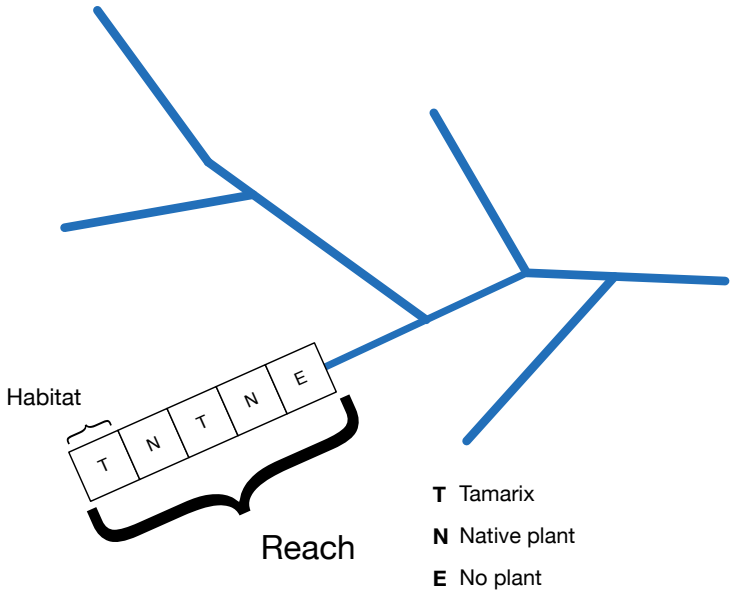


Figure 4.1: A river network, as modelled by the invasive species reinforcement learning environment.

Environment parameters The invasive species environment requires a number of parameters. For further explanation of the environment parameters consult the environment webpage[1]. The specific parameters used are presented in appendix A.

Agent Parameters The agents evaluated required different types of parameters. Some preliminary tests were ran and then the parameters giving best results were chosen. Parameters for MBIE are found in table 4.1.

Table 4.1: Parameters for MBIE

(a) Proper MBIE

Parameter	Value
Discount factor	0.9
Confidence	95%
$\Delta\omega$	$\frac{1}{2}\sqrt{\frac{2\lvert ln(2^{\lvert S\rvert}-2)-ln\delta\rvert}{N(s,a)}}$

(b) Realistic MBIE

Parameter	Value
Discount factor	0.9
$\Delta\omega$	$1-0.05N(s,a)$

[1] `http://2013.rl-competition.org/domains/invasive-species`

For DBN-E^3 a higher exploration limit resulted in the agent starting to exploit earlier but with a slightly lower final return. A higher partial state known limit resulted in later exploitation but no appreciable difference in final return. The values in table 4.2 were a good middle ground.

Table 4.2: DBN-E^3 parameters

Parameter	Value
Discount factor	0.9
Exploration limit	5%
Partial state known limit	5

Experiment parameters The tests performed had to be long enough to sample enough data to extract relevant results without making the running time too long. A single test consisted of a specific number of episodes with a specific length. A good combination was required to efficiently evaluate the agents. If a single episode consisted of too many samples it would be difficult to see the learning process as results are reported as total reward over an episode and that process might be hidden as its impact on the total reward is smaller with a longer episode. On the other hand, if an episode length is too short it would end before the agents could do any valuable learning.

In addition to a satisfactory episode length, a reasonable number of episodes needs to be sampled for it to be possible to draw conclusions from the results. If the number of episodes is too small the convergence of the agents cannot be seen. However, if the number of episodes are too large a lot of sampled data would be redundant for the study. Some preliminary experiments were run in order to tune the experiment parameters to suitable values. The episode length was set to 100 samples and there were 100 episodes per test.

To evaluate the agents in the invasive species environment combinations of reaches and number of habitats per reach that can be seen in table 4.3 were choosen. Combinations were chosen to have a wide range in the total number of states for the agents to deal with and to test how the agents deal with taking actions that have to take into account several state components.

Table 4.3: Combinations of reaches and habitats used in testing.

Reaches	Habitats	Total states
5	1	243
3	2	729
3	3	19 683
10	1	59 049
4	3	531 441
5	3	14 348 907

4.4 Programming environment

The Java programming language 1.7 was used to implement the agents. In addition, the java version of RL-Glue framework version 3.0 was used. For version control Git was used, due to its simplicity and performance.

4.5 RL-Glue

To evaluate the agents the RL-Glue framework was use, which acts as an interface for communication between the agent and the environment. The software uses the RL-Glue protocol, which specifies how a reinforcement learning problem should be divided when constructing experiments and how the different programs should communicate (Brian Tanner and Adam White 2009).

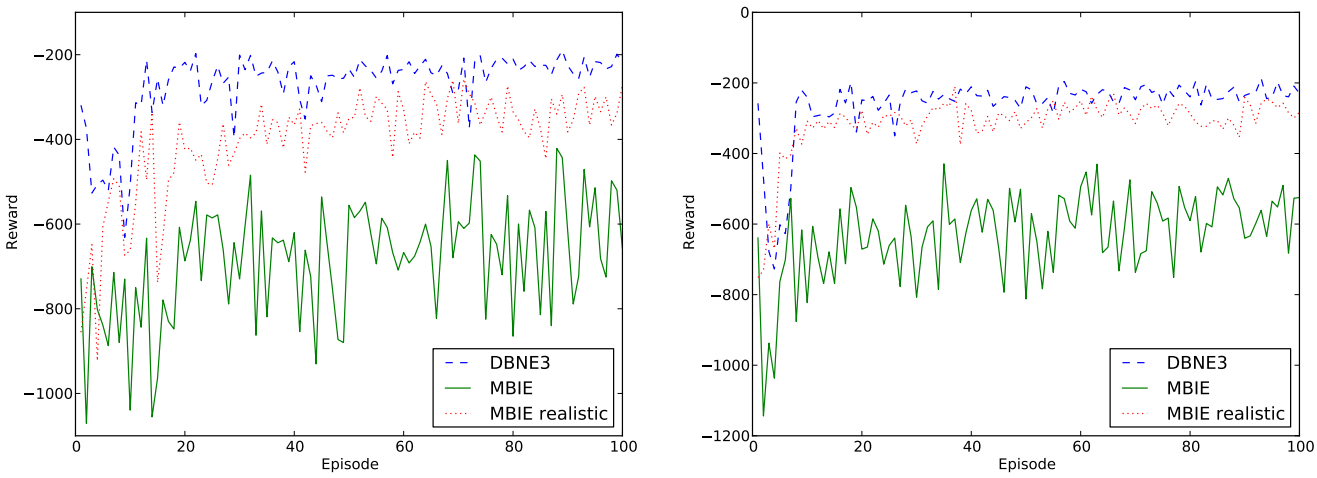
RL-Glue divides the reinforcement learning process into three separate programs: an agent, an environment and an experiment. RL-Glue provides a server software that manages the communication between these programs. The agent and the environment programs are responsible for executing the tasks as specified by RL-Glue and the experiment program acts as a bridge between the agent and environment (Brian Tanner and Adam White 2009).

The modular structure of RL-Glue makes it easier to construct repeatable reinforcement learning experiments. By separating the agent from the environment it is possible to reuse the environment and switch out the agent. It also makes it a lot easier to cooperate and continue working on existing environments implemented by other programmers, saving a lot of time.

Chapter 5

Results

In figures 5.1, 5.2 and 5.3 we have test results from running our agents on the invasive species environment on different sizes of river networks. The raw data can be found in appendix C.



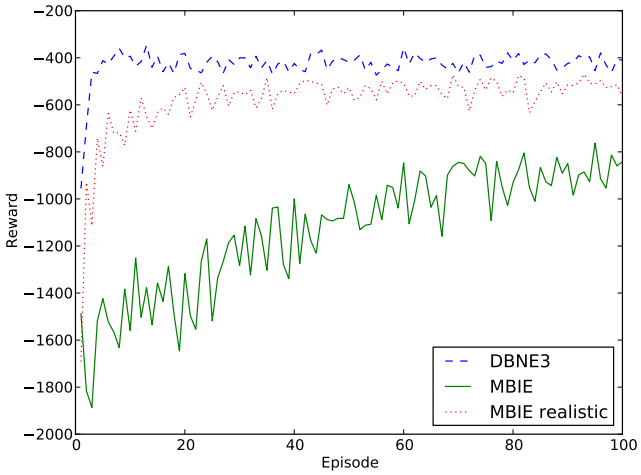
(a) 3 reaches and 3 habitats per reach

(b) 3 reaches and 2 habitats per reach

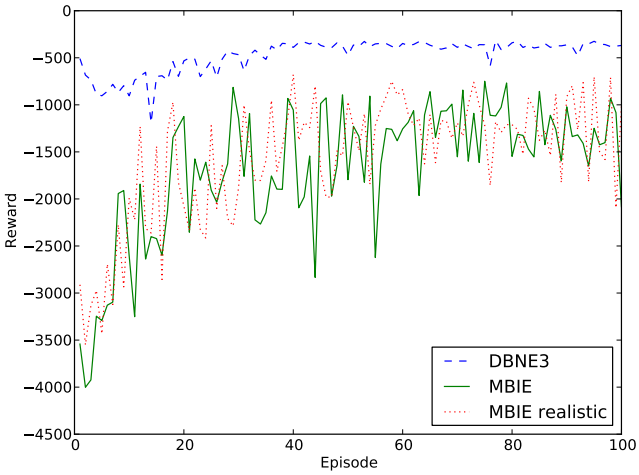
Figure 5.1: Test runs with different number of reaches and habitats

The agent learns for 100 episodes with 100 samples per episode, other parameters are specified in section 4.3. The invasive species environment associates a certain cost with each state and action, thus the reward is always negative. In each test the reward varies over time for the MBIE agent, the realistic MBIE agent and the DBN-E^3 agent.

With smaller state spaces (see figure 5.1a, 5.1b and 5.2a) we can see that

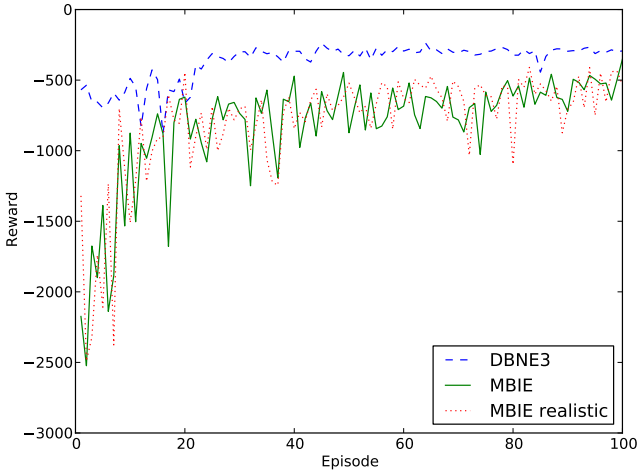


(a) 5 reaches and 1 habitats per reach

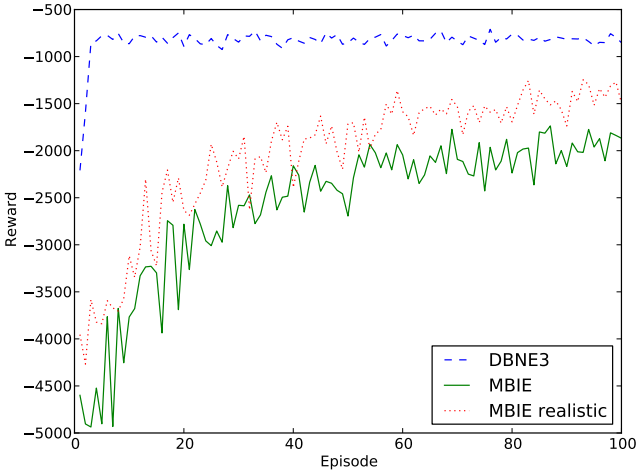


(b) 5 reaches and 3 habitats per reach

Figure 5.2: Test runs with different number of reaches and habitats



(a) 4 reaches and 3 habitats per reach



(b) 10 reaches and 1 habitats per reach

Figure 5.3: Test runs with different number of reaches and habitats

the realistic MBIE agent outperforms the proper MBIE agent and comes close to the DBN-E^3 agent. In the tests where the state space is larger (see figure 5.2b, 5.3b and 5.3a) proper MBIE and realistic MBIE are very close to each

other in performance while the DBN-E^3 agent outperforms both. The realistic MBIE algorithm performs better in with smaller state spaces due to more states becoming known quicker and thus having their true value.

In each of the test runs the DBN-E^3 algorithm exhibits a period of learning that corresponds to exploration (as described in section 3.2.1) where the agent has not explored the environment enough and seeks more information. This is apparent in figure 5.1a.

Chapter 6

Discussion B

This chapter is a discussion of the results and method used in the project. The algorithms are evaluated with regard to their performance achieved in the tests. Their strengths and weaknesses are discussed as well as their suitability for problems with an environment that has a large discrete state space. There is also a discussion on ethical aspects regarding the work in this thesis.

6.1 Evaluation of the agents

As an initial comment, the fact that we have reasonable results proves that the techniques used work. The tests completed within a reasonable time frame and did not run into hardware limitations. Furthermore, the results recorded show an increase in reward over time as expected of a working agent during its learning process.

6.1.1 E^3 in Factored Markov Decision Processes

In a comparison of the DBN-E^3 agent's behavior in the different-size environments, the similarity of the shape of the graphs is striking. At first there is a period of lower and lower performance. For the smallest environments, this period is very short, however. Next, there is a period of fairly constant high performance, which lasts until the end of the experiment. This behaviour is clearly visible in figures 5.1.

The similar shapes can be explained as a consequence of the different phases of the DBN-E^3 algorithm and the structure of the studied MDP. In the beginning of the experiment, the algorithm will spend almost all of its time in the balanced wandering and exploration phases. The longer the agent has been exploring, and the more states become known, the further the agent has to explore into hard-to-reach parts of the MDP to find unexplored states. Now, in the case of the invasive species environment with the parameters chosen as in the experiments presented, the most easily reachable states are the ones where there is

no tamarix infection. This means that the harder a state is to arrive at, the more infected reaches it will probably contain, and thus the performance of the DBN-E^3 agent in the exploration phase will fall as the experiment progresses. However, once the agent knows enough about the environment to enter the exploitation phase, the DBN-E^3 agent spends close to no time at all exploring unknown states, and it retains high performance.

Consequences of pooling observation data An interesting result is that the agent is able to enter the exploitation phase for the 10 reaches/1 habitat per reach test (see figure 5.3b) setup slightly earlier than for the 5 reaches/1 habitat per reach (see figure 5.2a) setup. This is probably connected to the optimization described in the last paragraph of section 3.2.2. In the 10 reaches/1 habitat per reach case there are nine reaches with two parents (including the reach itself) and one reach with only itself as parent. In the 5 reaches/1 habitat per reach case there are four reaches with two parents and one reach with only itself as parent. Since observations for state variables with similar parent structure are pooled by our DBN-E^3 implementation, the agent will receive almost double the information per time step for the two-parent reaches in the 10 reaches case as compared to the 5 reaches case.

Possible issues with the one-policy-per-reach optimization In section 3.2.3 an optimization is described that works very well for the particular environment and environment settings that the agent was tested for. However, this optimization makes several assumptions that may cause problems if the settings are changed. For instance, one assumption is that the state of a reach is only affected directly by its adjacent parents in river network. If the state of a reach was made to depend significantly on other reaches two or more levels up in the river network, the agent would probably not be able to converge on an optimal policy.

Another assumption that could lead to problems with other environment settings is the assumption that the maximal action cost in the environment is impossible or very hard to break. The invasive species environment has a maximum cost for actions. However, with the standard settings it is mathematically impossible to break this maximum. Our implementation of DBN-E^3 would achieve very poor performance if this was not the case, since a large penalty is given when the maximum action cost is breached.

DBN structure Finally, in the invasive species environment, the structure of the DBN underlying the MDP is known at the start of the experiment, so the agent does not need to infer it from its observations. If this was not the case, all the DBN optimizations would be useless unless some kind of algorithm for infering the DBN structure was added to the agent.

6.1.2 MBIE N

In comparison to the DBN-E^3 performance graphs, the MBIE performance exhibits a much smoother transition from poor to good performance. This is due to the fact that MBIE does not have a clear distinction between exploration and exploitation in phases. Instead, MBIE in effect always gives state-action pairs that are relatively unexplored a bonus to their expected value in order to promote exploration.

In the graphs for MBIE there are several "dips" in performance as for example in figure 5.2b. These could be explained as cases when the algorithm by chance enters previously unexplored states and spends several steps exploring this and similar/adjacent states.

Realistic MBIE and original MBIE The realistic version of MBIE outperforms the original MBIE in every test. This is probably explained by the fact that the state that is the easiest to arrive at is the one that gives the greatest reward (see section 6.1.1). Since the realistic version of MBIE considers states known and thus evaluates them realistically rather than optimistically much sooner than the original version of MBIE, it spends much less time exploring unknown states, which are bound to give lower rewards than the easier-explored states.

6.1.3 DBN-E^3 vs MBIE N

Perhaps belongs in conclusion?

An expectation on both agents is that they should converge on optimal behavior as t goes to infinity. However, it is clear from the results presented in chapter 5 that neither version of the MBIE agent reaches the same level of performance as the DBN-E^3 agent.

Unfair comparisons In one sense, the comparison between our implementations of MBIE and DBN-E^3 are not very fair. The DBN-E^3 implementation has been heavily optimized to work with factored MDPs and the invasive species environment in particular, whereas the MBIE implementation is much more generalized.

Large state spaces Stora pojkar

6.2 Potential factors impacting the results N

This sections is designated to discuss potential factors impacting the results collected in this thesis.

6.2.1 Impact of using one environment N

A pressing issue with the results presented in this report is that they are only collected from one environment. Due to limitations of the thesis only one environment has been when testing the algorithms. This presents a practical difficulty regarding the problem statement, considering that it is impossible evaluate the generality of the algorithms implemented and tested in this study. From the the results of the DBN-E^3 algorithm there are indications that the optimizations applied to the algorithm may be the specific for the Invasive Species environment. Which reduces the creditability regarding the generality conclusions presented in this thesis. The issue presented itself in the fact there is no way within the scope of this thesis to evaluate the performance of the agent without the optimizations applied.

Nevertheless, as discussed in section 6.1.1 the DBN-E^3 algorithm converge quick to an optimal policy, which takes a considerate longer time for the MBIE algorithm. Even though developed in parallel the generality of MBIE is more credible due to the fact it was simultaneously tested alongside Invasive Species in a simpler environment representing a grid world. Forcing the generality during construction of the agent. However even though the generality was partly forced during development, the lack of verification is hard to look over.

6.2.2 Implementation of algorithms N

One of the biggest challenges when constructing and evaluating algorithms is to validate the actual implementation of the algorithm and in addition there is no key to compare the result against. When building upon the work of another creator there is always a possibilities that specifications were interpreted poorly or mistakes during the implementation lead to less efficient or even wrong solution. As mentioned earlier in this paragraph, due to the lack of similar work containing results it is also hard to get an estimate of how well our algorithms actually perform in comparison with what they actually could have achieved.

In order to increase the credibility of the implementations used one method is to perform unit testing of the implementation. By creating implementations which enables unit testing it becomes easier to test the individual pieces of the algorithms and thereby verify correct behaviour. For more regarding information the implementation process see appendix B. Compare that process of using automated tests with manually validating the results of the complete algorithms behaviour as described in section 4.1, the advantages is obvious.

Kanske inrelevant, kanske inte - kanske behöver referense

6.2.3 Evaluation of experiments N

The evaluation of the algorithms is critical for achieving good results and as well deriving conslusions regarding the results collected. Evaluating agents is a tricky process due to the process of chosing approximatly correct parameters for both the algorithm and the environment. To be more specific for example it

is uncertain how much the paramaters for the MBIE affect the outcome of the experiment and trying out every diffrent combination is not possible within the scope of this theses. However, a possible solution for this problem is to derive the optimal paramaters using mathmatical proofs.

The same problem with the existed when chosing the problems for the environment, Invasive Species. The obvious problem is if the paramaters were chosen diffrently would the DBN-E^3 algorithm perform in the same way or if the optimizations discussed in section 6.1.1 would not perform in another possible setup. However, the MBIE algorithm there is a greater probability the results will be comparable to the results collected and discussed earlier due to the disussion in 6.2.1.

The last potential issue of the evaluation is a rather complex issue. Is the solution solution the agents finding, an optimal solution for the problem environment? When the number of habitans and reaches increases it becomes impossible to check manually of the policy computed by the algorithm is correct. As Dietterich, Taleghan, and Crowley (2013) mentions in their report as well as they are uncertain regarding the fact that their implementation achieves an optimal solution and should be taken into consideration when studying the policies used by the agents in this theis.

6.3 Similar studies N, K

The work with reducing the complexity of the reinforcement learning problems with environments containing large state spaces is not a new research topic. However, this thesis is taking a rather uncommon approach to the problem. Instead of trying to derive new algorithms for solving the problem, the focus is instead focused on usefulness of research conducted within the area of reinforcement learning problems. To be even more specific this thesis has zoomed in on two possible attempts on solving the problem with environments containing large state spaces. An approach that is rather uncommon and thus similar studies is as well uncommon. Thereby this section will combining focus on both studies similar to this thesis comparing the usefulness of research already conducted along with a summarized study of studies of work being done to reduce the complexity of environments with large state spaces.

Samma miljö kanske, dess pac?

Similar Studies Regarding Comparing Existing Research Borde finnas återkoppling i diskussion krin

It has been shown in (Strehl and Littman 2004) that MBIE outperforms E^3 in the MDPs RiverSwim and SixArms. (Dietterich, Taleghan, and Crowley 2013)

Large State Spaces using Dynamic Bayesian Networks This paragraph is devoted to similar methods using dynamic Bayesian networks as an underlying representation in order to model the structure of the environment. Due to the

relevance of using dynamic Bayesian network as a method for tackling similar problems it is thesis on its own in order to summarize them all. It is also possible to study section 7.2.2 for further references, however their main focus on the planning algorithm.

An algorithm to DBN-E^3 is the algorithm by Ross and Pineau (2012) which is as well utilizing a factored representation for the underlying structure. However, in the DBN-E^3 algorithm no planning algorithm was specified and therefore leaving an empty square in the implementation affecting the overall complexity of the algorithm. In the algorithm by Ross and Pineau (2012)

Resultatjämförelse

Similar Studies Regarding Large State Spaces Using Model Based Something grejen Grattis MBIE, er punkt.

6.4 Ethical aspects of Artificial Intelligence N ,K

What we do in this thesis is evaluating and comparing two already existing algorithms and techniques. Due to the fact that they both operate using simulations in this study it becomes hard to make any conclusions regarding its direct impact on matters such as ethics, social or economics on today's society. Therefore this sections focuses on some of the futher impact of reinforcemnet learning using models of the real world and simulations along with a high-level discussion regarding right or wrong with artificial intelligence.

6.4.1 Using models for simulating real world problems B,K

The environment Invasive Species Environment was used throughout this thesis. It is a simulation on the problem with invasive species and was first introduced in section 1.3. This domain focuses on the problem were the spreading process need to be controlled in a river network with native and invading plant species (Taleghanand, Crowley, and Dietterich 2014).

It is commonly known how fragile ecosystems is to changes. It is a complex system were one change could help the system short term while damage it long term. många actions Therefore using simulations with self learning algorithms allows to test more methods than time and money would allow. The simulations is a rough model of the real world so it's hard to capture all elements of the real world problem, therefore will the answers from the simulation also be roughly correct.

This presents a practical use of Reinforcement Learning which simplified can be viewed as a smart trial and error algorithm for finding optimal plans for working with real world problems like invasive species. This could revolutionize

the process of finding strategies for planning problems. For example let a computer use reinforcement learning algorithms to simulate different treatments for medicine and try to optimize the treatment without risking lives. This way of using artificial intelligence is discussed in the next section(6.4.2).

Se till att detta stycke kopplas hårt till efterföljande section

6.4.2 Artificial Intelligence, right or wrong? N

The restrictions on the possibilities of the applications of artificial intelligence is nonexistent. In a recent article by Stephen Hawking he raises a warning flag that we may are about the lose control of the world as we know it if we continue down the current path with the research in artificial intelligence. In may seem like the plot of a movie, however the author raises several worries that the machines will be able to outsmart humans in all areas from finance to military industry. He also points out that it should be noted that there is also several advantages and positive aspects of artificial intelligence (Hawking 2014).

One of these research areas is the field of autonomous cars. Google reported in 2014, that their research project with autonomous cars had covered 700 000 miles without human intervention (Urmson 2014). The key concept of autonomous cars cars is a safer traffic system, a system where the human factor is no longer is a part of the equation. Nevertheless, the question remains is the human factor really taken out of the equation? For the end user of the cars the answer is obviously yes. However, the human factor has been transformed into a different role and now lies on the engineers constructing the system for the car, making it able to drive without a driver. The need for worry is no longer for people driving under influence or being stressed or tired influencing their driving. Instead the human factor now lies on the people responsible for creating and deploying the system. Developing software is hard, it requires discipline and experience in order to avoid easy mistakes and errors. Nonetheless, the stakes are higher than when usually developing large scale systems, this time the systems were developed to be used by humans.

Moving the human factor, presents another relevant question lacking a good answer. Imagine a scenario where a collision is unavoidable for the computer driving the car. Lin (2014) raises several intresting aspects and scenarios in the article. However, the most releveant is a realistic but uncommon scenario, which vehicle should the car collide with. Forcing the ethical question onto the people responsible for the system. A choice no human would like to be responsible and in a way turning the driver-less vehicle into target seeking drone. Not every thing becomes better with technology, looking at the opposite case the driver properly do not have the time to think about the situation leaving the outcome to a sheer coincidence.

Chapter 7

Conclusion I

#winning

7.1 Last remarks I

7.2 Further work Ny text

As a result of the discussion carried out in chapter 6, the group has created this section containing some interesting issues with this thesis that we concentrated here as further work.

Se till att balansen DBN-E3 och MBIE är jämnare, enbart DBN-E3 nu

7.2.1 Testing algorithms in more environments N

As mentioned in the limitations (1.2) only one environment is used for collecting the results. Thereby, it is hard to verify how generic the algorithms performs. Due to the close ties to the environment Invasive Species during the development there is a risk that the algorithms does not perform well in other environments. Place has been left in the thesis for improvement and a possibility to conclude the generality of the algorithms studied in this thesis. One possible environment to extend the study with is the Tetris domain representing the game of Tetris from the 2008 edition of the Reinforcement Learning competition (Whiteson, B Tanner, and A White 2010).

Felaktig användning av källa? Hm...

7.2.2 Better Planning Algorithm for Factored E^3 in Markov Decision Processes N,K

In order to being able to completely take advantage of the factored structure, the complete process needs to be using factored versions. DBN-E^3 is utilizing

a dynamic Bayesian Network in order to factor the representation of the environment. In section 3.2.2, there were a discussion regarding which planning algorithm to use. In the end a slightly modified version of Value Iteration was used, due to its simplicity and given the scope of the project. However, as mentioned in the beginning of this section, the entire process from representation to planning needs utilize factored versions of the algorithm in order to maximize the output. Before the decision to use the modified Value Iteration, there were a discussion among alternative strategies to use in order to create an factored(approximate) value function and the two main options going forward are briefly presented below.

Approximate Value Determination Utilises a value determination algorithm to optimally approximate a value function, for factored representations using dynamic Bayesian Networks. The algorithm is using linear programming in order to achieve as good approximation as possible, over the factors associated with small subsets of problem features. However, as the authors of the algorithm mentions, their algorithm does not take advantage of the structured CPTs. Which leaves room for further improvements using dynamic programming steps (Koller and Parr 1999). Therefore it could be a good start when starting to improve the DBN-E^3 algorithm used in this thesis and step by step improve its efficiency.

Approximate Value Function Takes a different approach than the others options discussed, but the it attempts to solve the same problem, planning in large state spaces. This method represents the approximation of the value functions as a linear combination of basis functions. Each basis involves a small subset of the environment variables. A strength is that the algorithm comes in both an linear and dynamic programming versions. It could be more complex than the second option to implement, however Guestrin et al. (2003) presents results for problems with 10^{40} states. Which may very well result in a bigger improvement than the previous options discussed.

7.2.3 Improvements to the MBIE N

One extension that would be interesting to implement is the DDV-algorithm on which the author says: "(a) DDV requires fewer samples than existing methods to achieve good performance and (b) DDV terminates with a policy that is approximately optimal with high probability after only polynomially-many calls to the simulator" (Dietterich, Taleghan, and Crowley 2013).

Bibliography

Altman, Eitan (2002). "Applications of Markov decision processes in communication networks". In: *Handbook of Markov decision processes*. Springer, pp. 489–536.

Barto, Andrew G and Richard S. Sutton (1998). *Reinforcement learning: An introduction*. MIT press.

Bellman, Richard (1957). "A Markovian decision process". In: *Journal of Mathematics and Mechanics* 6.5, pp. 679–684.

Boutilier, Craig, Thomas Dean, and Steve Hanks (1999). "Decision-theoretic planning: Structural assumptions and computational leverage". In: *Journal Of Artificial Intelligence Research* 11, pp. 1–94.

Dietterich, Thomas, Majid Alkaee Taleghan, and Mark Crowley (2013). "PAC Optimal Planning for Invasive Species Management: Improved Exploration for Reinforcement Learning from Simulator-Defined MDPs". In: *Twenty-Seventh AAAI Conference on Artificial Intelligence*.

Gamma, Erich et al. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley Professional.

Good, Irving J (1953). "The population frequencies of species and the estimation of population parameters". In: *Biometrika* 40.3-4, pp. 237–264.

Guestrin, Carlos et al. (2003). "Efficient solution algorithms for factored MDPs". In: *J. Artif. Intell. Res.(JAIR)* 19, pp. 399–468.

Hawking, Stephen (2014). *Stephen Hawking: 'Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?'* URL: `http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html` (visited on 05/10/2014).

Heckerman, David (1998). "A Tutorial on Learning with Bayesian Networks". In: *Learning in Graphical Models*. Vol. 89. Springer Netherlands, pp. 301–354.

Kearns, Michael and Daphne Koller (1999). "Efficient reinforcement learning in factored MDPs". In: *IJCAI*. Vol. 16, pp. 740–747.

Kearns, Michael and Satinder Singh (2002). "Near-optimal reinforcement learning in polynomial time". In: *Machine Learning* 49.2-3, pp. 209–232.

Koller, Daphne and Ronald Parr (1999). "Computing factored value functions for policies in structured MDPs". In: *IJCAI*. Vol. 99, pp. 1332–1339.

Lin, Patrick (2014). *The Robot Car of Tomorrow May Just Be Programmed to Hit You*. URL: http://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/ (visited on 05/12/2014).

McCarthy, John (2007). *What is Artificial Intelligence?* URL: http://www-formal.stanford.edu/jmc/whatisai/ (visited on 02/10/2014).

Ng, Andrew Y et al. (2006). "Autonomous inverted helicopter flight via reinforcement learning". In: *Experimental Robotics IX*. Springer, pp. 363–372.

Ross, Stéphane and Joelle Pineau (2012). "Model-based Bayesian reinforcement learning in large structured domains". In: *arXiv preprint arXiv:1206.3281*.

Strehl, Alexander L and Michael L Littman (2004). "An empirical evaluation of interval estimation for markov decision processes". In: *Tools with Artificial Intelligence, 2004. ICTAI 2004. 16th IEEE International Conference on*. IEEE, pp. 128–135.

— (2008). "An analysis of model-based interval estimation for Markov decision processes". In: *Journal of Computer and System Sciences* 74.8, pp. 1309–1331.

Szepesvári, Csaba (2010). "Algorithms for reinforcement learning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 4.1, pp. 1–103.

Taleghanand, Majid A, Mark Crowley, and Thomas Dietterich (2014). *Invasive Species*. URL: https://sites.google.com/site/rlcompetition2014/domains/invasive-species (visited on 03/31/2014).

Tanner, Brian and Adam White (2009). "RL-Glue: Language-independent software for reinforcement-learning experiments". In: *The Journal of Machine Learning Research* 10, pp. 2133–2136.

Urmson, Chris (2014). *The latest chapter for the self-driving car: mastering city street driving*. URL: http://googleblog.blogspot.se/2014/04/the-latest-chapter-for-self-driving-car.html (visited on 05/10/2014).

Whiteson, S, B Tanner, and A White (2010). "The Reinforcement Learning Competitions". In: *AI-Magazine* 31.2, pp. 81–94.

Appendices

Appendix A

Environment specification

Table A.1: Dynamic parameters common to both species

Parameter	Value
Eradication rate	0.85
Restoration rate	0.65
Downstream spread rate	0.5
Upstream spread rate	0.1

Table A.2: Dynamic parameters different between the species

Parameter	Native	Tamarisk
Death rate	0.2	0.2
Production rate	200	200
Exogenous arrival[1]	Yes	Yes
Exogenous arrival probability	0.1	0.1
Exogenous arrival number	150	150

Table A.3: Cost function parameters

Parameter	Value
Cost per invaded reach	10
Cost per tree	0.1
Cost per empty slot	0.01
Eradication cost	0.5
Restoration cost	0.9

Table A.4: Variable costs depending on number of habitats affected by action

Parameter	Value
Eradication cost	0.4
Restoration cost for empty slot	0.4
Restoration cost for invaded slot	0.8

Appendix B

Comments on implementation process and tools

This appendix is designated for reflection and evaluation with regard to the processes or methods used throughout this thesis. Along with an evaluation regarding the framework or tool used and its impact on the work of the project.

B.1 Constructing Agents

Due to the complexity of testing the implementations it is hard to directly find the source of the faulty behaviour. In this project a lot of time has been spent tracking down errors in the code and two possible reasons for this problem has been identified.

The first bottleneck for the project was that during the implementation phase some of the details regarding the algorithm was unclear. When the errors in the code started appear it became hard to track down were they had occurred. If it would have been a logical error, the error would have been easier to track down, by stepping through the code step by step. Although, when not understanding the algorithm it becomes harder, a lot of time were spent trying to find out if it was a logical error or if the problem was that we implemented the algorithm wrong. Even though the implementation phase started with creating a high level description some details were left unclear, it still proved helpful during the high level reasoning going back and forward, trying to figure out the next step and what could be wrong with the current implementation.

The second reason is due to the fact that there were a problem finding a proper structure and design for the code. Because the lack of structure in the RL-Glue(B.3) library. The lacking structure of the code made it harder to test independent pieces of the algorithms and validate the correctness of the

pieces. It was not until the end of the phase the code started to become more structured and therefor it was easier to verify individual pieces of the code. The main reason to the lacking quality of the code is the rapid prototyping used when creating the agents, resulting in a working solution but maybe not the best way to do it. One might think that due to the small code base used in the project, the structure and design of the code is not as relevant as in a larger project. However, the structure and design still plays an important part and makes it easier to identify errors and simple mistakes.

B.2 Programming Language

To implement the algorithms the Java programming language was used as stated in section 4.4. A common rule when developing any kind of software is to use the right tool for the job, which is also a frequent comment when discussing which is the best programming language. Java is great for larger projects due to the fact that it is a statically typed and a object oriented programming language. Nonetheless, it also required a lot of boilerplate code for achieving the smallest of results.

For this reason it would be more suitable to use a dynamic programing language, for example Python. Using a more dynamic language is more suitable for rapid prototyping and experimenting in a way that Java can never be. Throughout the project the confidence in the choice of programming language decreased. The lack of support to easily implement the data structures needed for this project, lead to increased complexity in the code base and more error prone code. Instead of focusing on the core algorithm, focus was shifted towards trying to work around the problem with data structures and trying to create easier abstractions.

A problem that presented itself in the DBN-E^3algorithm when implementing the conditional probability table. The problem is when indexing with multiple keys and that for every level another table needed to be implemented, creating a chained and complex structure. To clarify HashMaps[1] was used for the implementation, resulting in a nested structure with a new level of HashMaps for every level and quickly getting out of hand. This is just one of many examples from the code created during the implementation phase. Nonetheless, while the need for abstractions is obvious it may very well be the fact it will prove to equally cumbersome to implement in another language.

To conclude some of the data structures proves to be rather complex to implement easily and use and may not be at the fault of Java, but when solely implementing algorithms the language should be a tool and barely noticeable and not subject to discussion and extra work. A possible solution to the problem with data structures for the conditional probability table is to instead use a data structure resembling a relationship database might be a better match, nevertheless the same kind of behaviour is easy to recreate in a programming language.

[1]HashMap http://docs.oracle.com/javase/7/docs/api/java/util/HashMap.html

B.3 RL-Glue Framework

Although RL-Glue provides the end user with many features and offers an easy interface for creating reinforcement learning experiments there are drawbacks with the framework. One major problem with the Java version of the RL-Glue framework is the lack of examples and documentation of the framework. Documentation of how the classes and their methods are supposed to work and interact with each-other is more or less non-existing, drastically increasing the level of effort needed to start working with RL-Glue. We found that the best way to start using RL-Glue was to either extend existing examples or inspecting the source of RL-Glue trying to figure out how the authors thought when they constructed the framework.

Another problem with the structure of the framework is that it does not use conventions of the Java programming language[2]. By creating classes that are responsible for holding data or information about an specific state and not implementing comparisons between these classes. It may seem as a technicality to not use the conventions for the language, however by not fulfilling the contracts for when designing a class, the error that occur due to contract is time spent wasted tracking down. Therefore we used an variation of the Facade and Adapter programming patterns (Gamma et al. 1994) to work around the shortcomings of the framework in an manageable and structured way.

[2] http://www.oracle.com/technetwork/java/javase/documentation/codeconvtoc-136057.html

Appendix C

Result tables

Table C.1: The realistic MBIE agent with 10 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-2,640.85	-2,482.25	-2,218.55	-1,789.65	-1,914.25	-2,209.11	361.93
2	-1,169.75	-2,266.85	-1,613.25	-1,990.25	-920.95	-1,592.21	557.47
3	-903.45	-842.85	-736.75	-1,019.95	-795.75	-859.75	108.49
4	-913.65	-723.25	-746.15	-972.75	-807.55	-832.67	107.55
5	-748.45	-837.75	-654.15	-901.95	-748.55	-778.17	94.89
6	-841.25	-826.05	-709.15	-736.75	-760.35	-774.71	57.03
7	-719.85	-964.15	-767.35	-878.35	-760.35	-818.01	100.65
8	-844.55	-686.15	-711.55	-772.15	-760.35	-754.95	61.18
9	-861.35	-871.45	-860.95	-854.75	-819.35	-853.57	20.04
10	-1,058.65	-873.15	-852.35	-748.55	-783.95	-863.33	120.25
11	-726.55	-888.45	-801.95	-689.55	-807.55	-782.81	77.46
12	-814.25	-756.95	-757.15	-713.15	-842.95	-776.89	51.50
13	-665.85	-920.35	-792.55	-807.55	-783.95	-794.05	90.38
14	-772.05	-831.05	-782.35	-783.95	-724.95	-778.87	37.80
15	-736.65	-981.05	-779.15	-760.35	-972.75	-845.99	120.48
16	-868.15	-709.75	-768.95	-842.95	-783.95	-794.75	62.69
17	-888.25	-863.05	-983.75	-866.55	-689.55	-858.23	106.30
18	-827.65	-895.05	-841.35	-795.75	-630.55	-798.07	100.28
19	-745.05	-793.95	-721.75	-831.15	-654.15	-749.21	68.05
20	-1,033.25	-1,001.15	-816.15	-783.95	-819.35	-890.77	116.79
21	-797.35	-837.85	-747.75	-713.15	-736.75	-766.57	50.31
22	-778.85	-617.05	-933.35	-890.15	-819.35	-807.75	122.37
23	-938.85	-915.25	-783.15	-866.55	-819.35	-864.63	64.74
24	-957.45	-901.85	-828.75	-866.55	-807.55	-872.43	59.68
25	-758.55	-852.95	-800.35	-866.55	-760.35	-807.75	50.55
26	-745.05	-1,053.35	-898.75	-960.95	-748.55	-881.33	134.57
27	-964.15	-1,184.85	-743.75	-819.35	-913.75	-925.17	168.22
28	-884.95	-797.35	-652.55	-677.75	-819.35	-766.39	98.28
29	-797.35	-1,043.35	-757.15	-665.95	-772.15	-807.19	141.02
30	-826.05	-955.75	-837.35	-736.75	-878.35	-846.85	79.84
31	-1,013.05	-834.35	-640.75	-724.95	-689.55	-780.53	148.21
32	-1,009.75	-692.85	-653.35	-831.15	-736.75	-784.77	142.12
33	-851.35	-819.25	-578.55	-618.75	-772.15	-728.01	122.23
34	-987.85	-868.15	-804.35	-772.15	-724.95	-831.49	101.74
35	-773.75	-625.45	-886.95	-819.35	-760.35	-773.17	96.26
36	-932.15	-831.05	-734.35	-736.75	-701.35	-787.13	94.40
37	-893.45	-994.55	-827.95	-713.15	-913.75	-868.57	105.28
38	-1,011.35	-1,109.05	-688.75	-913.75	-831.15	-910.81	162.07
39	-761.95	-1,019.75	-639.95	-972.75	-713.15	-821.51	166.15

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-753.55	-975.85	-793.35	-795.75	-677.75	-799.25	109.64
41	-920.35	-873.15	-766.55	-831.15	-783.95	-835.03	63.34
42	-942.25	-861.25	-800.35	-925.55	-760.35	-857.95	78.31
43	-866.45	-746.85	-827.95	-842.95	-701.35	-797.11	69.92
44	-858.05	-849.55	-873.55	-854.75	-654.15	-818.01	92.04
45	-815.75	-964.15	-780.75	-972.75	-772.15	-861.11	99.39
46	-681.05	-748.45	-697.35	-831.15	-783.95	-748.39	61.75
47	-819.25	-874.85	-731.95	-665.95	-925.55	-803.51	105.21
48	-709.65	-868.15	-849.95	-654.15	-748.55	-766.09	91.48
49	-879.85	-822.55	-805.15	-901.95	-937.35	-869.37	55.02
50	-849.55	-903.45	-901.15	-996.35	-618.75	-853.85	141.68
51	-750.15	-881.65	-627.35	-913.75	-842.95	-803.17	115.86
52	-888.25	-960.75	-709.95	-819.35	-854.75	-846.61	92.53
53	-955.75	-885.05	-733.55	-937.35	-842.95	-870.93	88.69
54	-901.75	-895.05	-734.35	-795.75	-713.15	-808.01	87.95
55	-652.45	-989.45	-828.75	-819.35	-677.75	-793.55	135.65
56	-765.35	-960.75	-653.35	-689.55	-760.35	-765.87	118.84
57	-964.15	-911.95	-781.55	-831.15	-949.15	-887.59	78.54
58	-842.75	-964.15	-794.95	-842.95	-701.35	-829.23	95.00
59	-812.55	-900.15	-675.35	-748.55	-665.95	-760.51	98.16
60	-959.05	-852.95	-817.75	-689.55	-559.75	-775.81	154.44
61	-746.85	-888.35	-864.95	-654.15	-842.95	-799.45	97.42
62	-729.95	-740.05	-830.35	-772.15	-795.75	-773.65	41.06
63	-793.95	-903.45	-779.15	-972.75	-713.15	-832.49	104.02
64	-756.95	-869.75	-816.95	-901.95	-819.35	-832.99	55.53
65	-859.65	-871.45	-919.95	-701.35	-724.95	-815.47	96.46
66	-745.05	-1,007.95	-642.35	-748.55	-618.75	-752.53	154.39
67	-782.15	-814.15	-628.95	-677.75	-736.75	-727.95	75.44
68	-768.75	-1,024.95	-838.15	-736.75	-890.15	-851.75	113.77
69	-738.35	-810.85	-779.15	-724.95	-913.75	-793.41	75.35
70	-669.25	-982.75	-887.75	-783.95	-854.75	-835.69	117.35
71	-743.35	-986.05	-917.55	-831.15	-890.15	-873.65	91.68
72	-605.25	-809.15	-932.55	-795.75	-760.35	-780.61	117.57
73	-842.85	-864.75	-779.95	-689.55	-842.95	-804.01	71.40
74	-1,033.15	-829.35	-781.55	-819.35	-677.75	-828.23	129.32
75	-1,023.15	-891.75	-817.75	-807.55	-819.35	-871.91	90.97
76	-829.35	-748.35	-688.75	-654.15	-630.55	-710.23	80.00
77	-740.05	-1,055.25	-886.95	-713.15	-842.95	-847.67	136.32
78	-772.05	-814.25	-687.15	-866.55	-842.95	-796.59	70.60
79	-820.95	-893.35	-864.95	-677.75	-807.55	-812.91	82.98
80	-718.15	-1,011.25	-793.35	-654.15	-630.55	-761.49	153.25
81	-773.75	-802.45	-719.35	-772.15	-795.75	-772.69	32.65
82	-932.15	-805.75	-791.75	-724.95	-866.55	-824.23	78.57
83	-987.75	-790.65	-711.55	-783.95	-795.75	-813.93	103.04
84	-826.05	-844.55	-758.75	-890.15	-819.35	-827.77	47.46
85	-825.95	-675.95	-745.35	-890.15	-842.95	-796.07	85.06
86	-824.25	-1,095.65	-860.15	-701.35	-701.35	-836.55	161.55
87	-851.25	-986.05	-746.15	-819.35	-689.55	-818.47	112.93
88	-943.95	-954.15	-651.75	-748.55	-654.15	-790.51	149.94
89	-827.75	-788.85	-792.55	-878.35	-724.95	-802.49	56.33
90	-987.75	-992.85	-593.55	-642.35	-866.55	-816.61	189.05
91	-837.85	-915.35	-779.95	-819.35	-701.35	-810.77	78.53
92	-763.65	-858.05	-757.15	-807.55	-878.35	-812.95	54.51
93	-812.45	-733.35	-814.55	-831.15	-807.55	-799.81	38.20
94	-814.25	-708.05	-1,011.95	-842.95	-819.35	-839.31	109.63
95	-982.75	-937.25	-850.75	-866.55	-783.95	-884.25	77.47
96	-863.05	-927.05	-790.95	-878.35	-783.95	-848.67	60.73
97	-969.25	-986.05	-757.95	-854.75	-689.55	-851.51	129.38
98	-692.85	-824.25	-591.95	-878.35	-795.75	-756.63	114.13
99	-822.65	-911.95	-722.55	-748.55	-772.15	-795.57	74.78
100	-863.05	-859.75	-898.75	-831.15	-807.55	-852.05	34.56

Table C.2: The realistic MBIE agent with 3 reaches and 2 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-326.90	-270.50	-306.30	-435.60	-369.30	-257.80	63.48
2	-557.90	-162.40	-503.80	-651.60	-467.60	-468.66	184.70
3	-924.40	-498.20	-912.20	-624.60	-429.70	-677.82	230.43
4	-1,108.40	-322.30	-766.50	-756.60	-682.70	-727.30	280.02
5	-738.30	-578.80	-449.80	-714.40	-527.80	-601.82	122.90
6	-335.60	-608.00	-441.00	-780.80	-970.90	-627.26	255.83
7	-255.30	-256.10	-300.10	-847.10	-837.00	-499.12	313.60
8	-188.20	-197.50	-311.40	-281.50	-291.00	-253.92	56.88
9	-171.40	-241.90	-211.80	-223.20	-257.10	-221.08	32.75
10	-197.50	-266.30	-246.30	-234.40	-257.80	-240.46	26.86
11	-174.60	-423.20	-183.60	-294.30	-403.30	-295.80	117.33
12	-183.90	-471.40	-244.60	-232.10	-329.70	-292.34	113.05
13	-245.60	-256.90	-270.80	-170.00	-501.30	-288.92	124.98
14	-345.80	-187.30	-485.60	-265.20	-198.20	-296.42	123.24
15	-275.10	-221.10	-455.90	-306.40	-169.80	-285.66	108.53
16	-196.00	-187.90	-249.90	-246.30	-210.90	-218.20	28.54
17	-257.60	-255.60	-304.60	-242.20	-223.60	-256.72	30.01
18	-147.90	-245.60	-141.00	-268.20	-188.20	-198.18	57.12
19	-254.20	-241.60	-247.20	-682.60	-270.80	-339.28	192.23
20	-229.20	-255.60	-268.30	-255.70	-230.90	-247.94	17.14
21	-241.50	-284.40	-270.80	-285.30	-175.50	-251.50	46.03
22	-288.50	-234.50	-185.70	-247.20	-235.40	-238.26	36.70
23	-298.00	-289.10	-282.60	-258.10	-259.90	-277.54	17.80
24	-238.40	-269.00	-187.30	-281.00	-314.20	-257.98	47.94
25	-264.00	-336.20	-234.50	-198.20	-138.90	-234.36	73.60
26	-859.60	-220.20	-175.50	-259.90	-235.50	-350.14	286.45
27	-221.80	-303.30	-225.40	-273.50	-259.00	-256.60	34.12
28	-233.30	-241.90	-197.50	-208.40	-207.50	-217.72	18.89
29	-248.10	-199.10	-257.20	-235.40	-199.10	-227.78	27.30
30	-173.90	-221.80	-223.60	-237.20	-256.50	-222.60	30.56
31	-210.20	-257.40	-212.70	-291.00	-282.60	-250.78	37.98
32	-291.90	-222.70	-243.80	-293.50	-234.60	-257.30	33.17
33	-220.90	-209.90	-210.90	-256.50	-222.90	-224.22	18.96
34	-190.00	-247.20	-230.10	-246.30	-315.50	-245.82	45.33
35	-270.10	-185.10	-234.50	-184.80	-290.30	-232.96	48.17
36	-238.10	-186.60	-282.60	-281.90	-236.30	-245.10	39.72
37	-208.40	-188.20	-314.60	-327.80	-222.70	-252.34	64.21
38	-187.30	-234.50	-199.10	-267.60	-199.10	-217.52	33.12
39	-185.30	-332.50	-235.40	-185.70	-212.70	-230.32	60.82
40	-210.90	-188.20	-188.20	-220.60	-247.20	-211.02	24.71
41	-279.20	-161.20	-163.70	-330.70	-246.30	-236.22	73.76
42	-211.80	-223.60	-210.20	-305.30	-235.40	-237.26	39.37
43	-207.50	-219.50	-174.60	-247.20	-255.70	-220.90	32.51
44	-268.30	-247.20	-341.40	-277.80	-197.50	-266.44	52.14
45	-282.60	-222.00	-213.60	-306.20	-236.30	-252.14	40.30
46	-188.20	-210.90	-361.20	-176.40	-257.20	-238.78	75.09
47	-223.60	-176.40	-223.60	-291.70	-292.60	-241.58	50.03
48	-337.50	-267.40	-138.00	-259.20	-244.50	-249.32	71.79
49	-222.00	-259.00	-420.20	-187.30	-271.00	-271.90	89.17
50	-200.00	-247.20	-197.50	-271.70	-138.30	-210.94	51.44
51	-163.70	-247.20	-176.40	-314.60	-196.60	-219.70	61.86
52	-234.50	-267.40	-248.10	-247.20	-301.40	-259.72	26.10
53	-282.60	-244.90	-221.10	-249.00	-293.70	-258.26	29.55
54	-320.70	-211.80	-303.70	-173.70	-198.30	-241.64	66.12
55	-561.30	-165.50	-199.10	-235.40	-270.80	-286.42	158.63
56	-210.90	-267.60	-222.40	-188.20	-151.00	-208.02	43.04
57	-175.50	-281.00	-162.80	-163.70	-196.60	-195.92	49.48
58	-270.80	-247.20	-256.30	-210.90	-162.10	-229.46	43.66
59	-211.80	-260.80	-188.20	-273.50	-237.20	-234.30	34.92
60	-222.70	-233.80	-186.40	-199.10	-248.20	-218.04	25.20
61	-236.30	-258.10	-259.00	-161.20	-210.20	-224.96	40.83
62	-269.90	-259.00	-234.50	-199.10	-326.40	-257.78	47.02

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-317.10	-237.20	-188.20	-245.40	-173.00	-232.18	56.67
64	-130.10	-255.60	-220.20	-293.50	-196.60	-219.20	61.84
65	-139.20	-211.80	-222.70	-256.50	-291.00	-224.24	56.76
66	-255.60	-210.90	-293.70	-296.20	-211.80	-253.64	41.82
67	-185.70	-329.10	-244.70	-269.90	-259.90	-257.86	51.49
68	-231.30	-270.80	-176.40	-200.00	-187.30	-213.16	38.23
69	-269.00	-259.00	-196.60	-162.80	-212.70	-220.02	44.14
70	-270.80	-164.60	-248.10	-306.40	-249.00	-247.78	52.17
71	-224.50	-140.10	-255.60	-222.70	-211.00	-210.78	42.82
72	-187.30	-247.20	-175.50	-151.90	-259.90	-204.36	46.89
73	-233.80	-211.80	-257.40	-222.70	-210.00	-227.14	19.42
74	-199.10	-223.60	-140.10	-269.20	-269.10	-220.22	54.00
75	-256.30	-294.60	-283.50	-244.70	-223.60	-260.54	28.81
76	-210.00	-270.80	-199.10	-197.50	-210.90	-217.66	30.33
77	-140.10	-269.00	-302.80	-235.40	-244.70	-238.40	60.82
78	-150.30	-243.80	-190.00	-248.10	-199.10	-206.26	40.64
79	-209.30	-233.80	-258.10	-187.30	-224.50	-222.60	26.52
80	-235.40	-295.30	-256.30	-223.60	-187.30	-239.58	39.97
81	-184.80	-200.00	-211.80	-174.60	-210.10	-196.26	16.18
82	-233.80	-330.90	-235.40	-280.10	-233.80	-262.80	42.92
83	-209.10	-234.50	-200.90	-174.60	-199.20	-203.66	21.52
84	-274.40	-239.90	-256.50	-255.60	-200.00	-245.28	28.11
85	-280.80	-199.10	-285.30	-235.40	-232.80	-246.68	36.19
86	-232.00	-259.90	-270.80	-201.80	-257.30	-244.36	27.71
87	-173.00	-259.00	-210.90	-281.70	-200.00	-224.92	44.44
88	-233.60	-255.60	-174.80	-210.90	-186.50	-212.28	33.18
89	-223.60	-259.00	-282.60	-221.80	-281.70	-253.74	29.88
90	-259.90	-175.50	-210.90	-297.10	-233.60	-235.40	46.36
91	-249.00	-244.70	-256.50	-200.00	-198.50	-229.74	28.16
92	-232.90	-223.60	-221.80	-257.20	-199.10	-226.92	21.00
93	-200.90	-196.60	-141.00	-210.90	-186.60	-187.20	27.26
94	-211.80	-306.20	-259.20	-234.50	-210.90	-244.52	39.75
95	-316.40	-235.40	-186.40	-232.20	-233.60	-240.80	46.98
96	-175.50	-188.20	-175.50	-211.80	-236.30	-197.46	26.29
97	-294.40	-222.00	-221.80	-186.40	-255.60	-236.04	40.78
98	-245.60	-246.30	-234.50	-200.90	-270.00	-239.46	25.14
99	-200.00	-234.50	-210.00	-175.50	-224.50	-208.90	22.88
100	-232.90	-174.60	-247.20	-247.20	-246.40	-229.66	31.38

Table C.3: The realistic MBIE agent with 3 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-427.15	-257.25	-255.85	-321.95	-334.75	-319.39	70.27
2	-271.55	-408.95	-497.75	-407.45	-283.05	-373.75	95.42
3	-444.55	-432.55	-353.45	-617.65	-785.75	-526.79	173.93
4	-613.85	-499.55	-566.45	-456.75	-397.45	-506.81	85.93
5	-600.15	-547.45	-408.75	-511.05	-413.65	-496.21	83.84
6	-708.35	-388.05	-576.75	-439.35	-488.95	-520.29	126.07
7	-630.35	-160.95	-659.65	-462.05	-178.55	-418.31	239.19
8	-502.65	-178.75	-597.75	-656.25	-251.35	-437.35	211.77
9	-970.65	-342.85	-1,100.55	-547.25	-195.25	-631.31	392.35
10	-976.25	-258.05	-531.25	-623.15	-160.25	-509.79	322.66
11	-612.85	-248.65	-346.45	-185.65	-174.05	-313.53	180.76
12	-300.05	-301.05	-230.45	-455.15	-308.15	-318.97	82.43
13	-310.45	-209.65	-245.35	-112.45	-185.65	-212.71	73.16
14	-246.65	-266.35	-240.95	-674.25	-253.65	-336.37	189.12
15	-317.05	-184.95	-301.05	-231.65	-250.55	-257.05	53.44
16	-780.05	-175.35	-217.95	-242.75	-183.85	-319.99	258.60
17	-209.05	-234.55	-288.25	-252.75	-312.65	-259.45	41.43
18	-289.65	-205.25	-172.05	-233.75	-244.15	-228.97	43.98

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-217.35	-207.15	-319.45	-194.25	-221.25	-231.89	50.05
20	-161.55	-253.55	-260.75	-148.65	-265.75	-218.05	57.81
21	-216.85	-245.85	-253.15	-233.45	-242.15	-238.29	13.93
22	-196.65	-246.75	-162.95	-195.85	-184.75	-197.39	30.76
23	-317.05	-199.45	-388.95	-452.25	-242.95	-320.13	103.35
24	-207.45	-348.75	-376.15	-328.15	-279.25	-307.95	66.42
25	-277.95	-302.75	-344.75	-206.05	-183.15	-262.93	67.28
26	-125.85	-231.35	-302.75	-255.05	-218.35	-226.67	64.90
27	-230.95	-297.65	-257.65	-333.45	-217.65	-267.47	47.90
28	-301.75	-171.45	-256.65	-257.05	-277.55	-252.89	49.14
29	-257.35	-467.65	-374.25	-550.15	-331.75	-396.23	114.80
30	-195.65	-197.45	-208.15	-207.55	-195.65	-200.89	6.40
31	-207.45	-255.05	-241.55	-231.95	-241.45	-235.49	17.70
32	-148.45	-210.75	-231.95	-162.05	-255.15	-201.67	45.45
33	-195.65	-389.95	-265.15	-185.65	-217.55	-250.79	83.61
34	-221.95	-278.85	-287.85	-244.65	-183.85	-243.43	42.55
35	-162.25	-308.15	-245.55	-255.65	-237.35	-241.79	52.34
36	-252.45	-241.35	-185.65	-211.05	-183.45	-214.79	31.50
37	-244.25	-207.65	-209.25	-406.95	-136.75	-240.97	100.67
38	-255.85	-337.45	-315.85	-313.15	-239.05	-292.27	42.40
39	-162.05	-209.55	-258.25	-280.05	-232.15	-228.41	45.64
40	-173.55	-241.25	-185.65	-230.55	-252.05	-216.61	34.89
41	-242.75	-284.95	-269.15	-210.15	-524.35	-306.27	125.17
42	-255.25	-739.55	-219.25	-209.25	-337.45	-352.15	222.36
43	-255.55	-171.75	-282.95	-270.05	-268.85	-249.83	44.71
44	-267.75	-303.35	-209.25	-278.65	-301.25	-272.05	38.20
45	-244.65	-230.65	-212.85	-279.85	-586.95	-310.99	156.21
46	-267.95	-256.75	-229.15	-309.95	-185.65	-249.89	46.20
47	-267.55	-256.15	-221.95	-279.15	-217.85	-248.53	27.41
48	-230.45	-291.15	-264.95	-208.35	-289.35	-256.85	36.56
49	-227.55	-195.05	-221.95	-437.65	-196.55	-255.75	102.73
50	-333.95	-269.05	-172.95	-184.75	-209.25	-233.99	67.05
51	-242.55	-303.85	-242.45	-247.35	-232.85	-253.81	28.46
52	-196.85	-162.05	-291.35	-222.85	-184.95	-211.61	49.67
53	-171.75	-314.65	-190.45	-234.65	-221.15	-226.53	55.14
54	-241.95	-253.35	-213.05	-270.05	-206.95	-237.07	26.74
55	-278.65	-184.05	-158.05	-351.75	-303.45	-255.19	81.70
56	-206.35	-231.95	-254.55	-234.65	-267.35	-238.97	23.35
57	-197.45	-147.95	-236.75	-148.45	-279.45	-202.01	57.04
58	-256.45	-184.75	-245.95	-160.55	-496.85	-268.91	133.66
59	-209.05	-302.95	-257.85	-186.55	-229.25	-237.13	45.20
60	-231.95	-221.05	-226.45	-232.85	-262.05	-234.87	15.92
61	-197.75	-292.05	-225.85	-186.55	-181.05	-216.65	45.55
62	-287.75	-286.05	-199.65	-231.95	-216.95	-244.47	40.39
63	-195.65	-278.75	-255.35	-221.95	-181.65	-226.67	40.45
64	-257.35	-221.05	-225.05	-211.05	-139.45	-210.79	43.49
65	-269.35	-218.55	-226.55	-234.65	-264.35	-242.69	22.85
66	-290.25	-256.45	-305.95	-231.95	-160.45	-249.01	57.30
67	-162.95	-196.55	-289.15	-232.85	-249.25	-226.15	48.52
68	-247.35	-292.05	-245.55	-172.95	-264.35	-244.45	44.11
69	-210.15	-533.45	-261.25	-237.35	-214.65	-291.37	136.85
70	-210.15	-149.45	-272.25	-280.95	-499.15	-282.39	132.27
71	-162.95	-290.35	-218.15	-207.45	-160.25	-207.83	52.90
72	-256.65	-244.65	-227.15	-199.25	-938.45	-373.23	316.70
73	-184.75	-173.85	-203.95	-245.55	-267.15	-215.05	39.95
74	-210.15	-231.45	-269.85	-174.75	-126.65	-202.57	54.67
75	-255.55	-327.25	-276.35	-233.75	-233.45	-265.27	38.94
76	-211.05	-276.85	-225.65	-246.45	-199.25	-231.85	30.71
77	-198.35	-221.15	-192.85	-161.15	-282.15	-211.13	45.11
78	-232.85	-139.35	-226.65	-221.05	-221.05	-208.19	38.79
79	-242.85	-219.45	-224.95	-199.25	-221.05	-221.51	15.55
80	-245.55	-207.65	-229.15	-221.05	-149.35	-210.55	36.86
81	-135.95	-209.35	-350.15	-245.55	-244.65	-237.13	77.31
82	-244.65	-242.35	-202.05	-267.35	-185.65	-228.41	33.53
83	-256.45	-199.25	-226.55	-211.95	-174.75	-213.79	30.50

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-171.35	-208.35	-256.15	-197.45	-306.35	-227.93	53.53
85	-160.25	-187.45	-243.65	-275.45	-256.45	-224.65	48.69
86	-162.05	-220.15	-292.45	-243.05	-208.35	-225.21	47.80
87	-326.35	-208.45	-205.55	-232.55	-264.45	-247.47	50.03
88	-265.95	-196.65	-202.95	-232.85	-162.95	-212.27	38.95
89	-257.35	-139.35	-123.35	-221.05	-208.55	-189.93	56.68
90	-246.45	-196.75	-327.85	-136.65	-222.45	-226.03	70.08
91	-208.35	-301.85	-224.15	-256.65	-233.05	-244.81	36.36
92	-280.05	-221.05	-302.15	-267.65	-210.15	-256.21	39.27
93	-150.25	-184.75	-254.05	-221.05	-219.55	-205.93	39.62
94	-246.45	-347.15	-285.35	-189.25	-193.35	-252.31	66.28
95	-255.55	-257.45	-233.75	-207.45	-127.55	-216.35	53.60
96	-186.55	-243.05	-245.55	-195.65	-221.95	-218.55	26.88
97	-245.55	-174.75	-245.55	-232.85	-269.15	-233.57	35.40
98	-234.65	-232.85	-254.05	-232.85	-186.55	-228.19	24.94
99	-222.85	-172.95	-234.65	-235.55	-127.55	-198.71	47.33
100	-288.25	-230.35	-173.85	-136.65	-266.75	-219.17	63.29

Table C.4: The realistic MBIE agent with 4 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-680.30	-421.40	-580.00	-582.50	-580.00	-568.84	93.00
2	-421.50	-478.80	-575.30	-631.50	-575.30	-536.48	84.54
3	-608.90	-497.20	-744.90	-669.10	-744.90	-653.00	104.13
4	-1,035.50	-516.20	-494.10	-738.20	-494.10	-655.62	235.98
5	-556.40	-992.10	-781.50	-368.60	-781.50	-696.02	239.26
6	-830.10	-625.20	-668.90	-518.50	-668.90	-662.32	112.14
7	-574.90	-645.70	-660.60	-417.40	-660.60	-591.84	103.77
8	-838.50	-724.10	-593.00	-462.20	-593.00	-642.16	143.60
9	-832.00	-647.20	-485.10	-487.80	-485.10	-587.44	153.51
10	-758.00	-594.10	-388.80	-307.60	-388.80	-487.46	184.65
11	-476.40	-676.40	-463.00	-660.90	-463.00	-547.94	110.46
12	-696.50	-1,287.10	-843.80	-473.40	-843.80	-828.92	297.62
13	-866.30	-685.00	-350.40	-546.40	-350.40	-559.70	222.20
14	-292.00	-332.50	-452.40	-608.20	-452.40	-427.50	123.77
15	-573.40	-385.50	-544.50	-432.50	-544.50	-496.08	82.06
16	-416.10	-995.90	-1,169.70	-650.90	-1,169.70	-880.46	335.03
17	-547.10	-679.30	-615.10	-388.50	-615.10	-569.02	111.22
18	-713.10	-545.40	-536.70	-576.10	-536.70	-581.60	75.27
19	-480.30	-433.20	-498.30	-551.60	-498.30	-492.34	42.50
20	-847.90	-571.00	-540.60	-812.40	-540.60	-662.50	154.06
21	-1,118.30	-390.80	-518.70	-560.70	-518.70	-621.44	284.98
22	-243.60	-569.80	-243.10	-667.80	-243.10	-393.48	208.59
23	-433.70	-620.90	-359.40	-334.80	-359.40	-421.64	117.41
24	-322.60	-441.60	-253.00	-518.10	-253.00	-357.66	118.24
25	-308.60	-323.90	-365.90	-194.20	-365.90	-311.70	70.43
26	-270.50	-436.00	-312.40	-335.50	-312.40	-333.36	61.99
27	-276.00	-437.00	-371.40	-277.50	-371.40	-346.66	69.21
28	-310.80	-433.10	-347.10	-441.60	-347.10	-375.94	58.06
29	-251.40	-530.80	-294.60	-290.20	-294.60	-332.32	112.43
30	-284.50	-349.90	-289.30	-283.00	-289.30	-299.20	28.48
31	-305.60	-230.30	-312.30	-334.60	-312.30	-299.02	39.95
32	-287.20	-311.00	-221.20	-652.60	-221.20	-338.64	179.98
33	-309.20	-338.40	-207.60	-290.50	-207.60	-270.66	60.04
34	-379.20	-228.50	-278.40	-313.50	-278.40	-295.60	55.69
35	-354.30	-302.80	-338.30	-232.30	-338.30	-313.20	49.00
36	-311.50	-379.30	-254.80	-321.80	-254.80	-304.44	52.16
37	-276.80	-322.90	-345.80	-384.60	-345.80	-335.18	39.46
38	-354.00	-316.20	-317.60	-558.50	-317.60	-372.78	105.04
39	-265.50	-240.10	-312.80	-331.00	-312.80	-292.44	38.02

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-343.10	-273.90	-278.60	-303.80	-278.60	-295.60	29.04
41	-320.20	-253.70	-265.70	-372.80	-265.70	-295.62	50.27
42	-563.20	-206.90	-338.30	-302.10	-338.30	-349.76	130.86
43	-325.80	-305.60	-426.30	-376.40	-426.30	-372.08	55.81
44	-277.70	-299.00	-315.60	-268.40	-315.60	-295.26	21.63
45	-244.90	-265.00	-219.40	-244.80	-219.40	-238.70	19.44
46	-372.30	-216.90	-252.30	-275.20	-252.30	-273.80	58.88
47	-221.30	-334.90	-292.00	-347.10	-292.00	-297.46	49.31
48	-349.40	-356.40	-219.40	-255.70	-219.40	-280.06	68.17
49	-298.00	-243.10	-369.00	-315.60	-369.00	-318.94	52.95
50	-365.30	-303.00	-326.50	-315.60	-326.50	-327.38	23.31
51	-347.10	-287.70	-254.80	-305.60	-254.80	-290.00	38.69
52	-300.50	-254.80	-394.30	-302.00	-394.30	-329.18	62.40
53	-298.90	-316.50	-207.60	-327.10	-207.60	-271.54	59.23
54	-320.10	-289.50	-373.10	-418.40	-373.10	-354.84	50.46
55	-287.90	-334.90	-241.20	-266.60	-241.20	-274.36	39.08
56	-357.90	-290.30	-318.30	-335.80	-318.30	-324.12	24.95
57	-301.60	-288.60	-294.70	-242.10	-294.70	-284.34	24.06
58	-260.20	-353.70	-275.30	-327.40	-275.30	-298.38	40.08
59	-239.80	-264.10	-312.30	-184.90	-312.30	-262.68	53.62
60	-418.60	-302.30	-264.90	-221.20	-264.90	-294.38	75.14
61	-278.60	-324.70	-267.80	-266.60	-267.80	-281.10	24.86
62	-184.90	-344.30	-338.30	-325.60	-338.30	-306.28	68.20
63	-343.90	-335.00	-279.30	-267.50	-279.30	-301.00	35.57
64	-298.80	-240.20	-161.30	-325.30	-161.30	-237.38	75.97
65	-220.40	-330.20	-344.40	-230.30	-344.40	-293.94	62.98
66	-324.90	-220.40	-325.60	-231.20	-325.60	-285.54	54.67
67	-345.80	-372.10	-316.50	-207.80	-316.50	-311.74	62.55
68	-288.40	-267.50	-314.70	-207.60	-314.70	-278.58	44.35
69	-298.10	-253.00	-300.90	-359.60	-300.90	-302.50	37.87
70	-311.10	-346.40	-267.50	-343.50	-267.50	-307.20	38.80
71	-288.70	-325.70	-243.90	-339.20	-243.90	-288.28	44.53
72	-265.10	-218.00	-362.80	-281.10	-362.80	-297.96	63.57
73	-337.10	-383.00	-281.10	-329.20	-281.10	-322.30	42.85
74	-336.90	-335.80	-339.40	-233.00	-339.40	-316.90	46.93
75	-277.30	-291.20	-313.10	-292.00	-313.10	-297.34	15.53
76	-392.40	-265.80	-278.40	-256.60	-278.40	-294.32	55.59
77	-276.20	-276.90	-266.60	-300.40	-266.60	-277.34	13.82
78	-300.70	-218.00	-277.70	-266.80	-277.70	-268.18	30.65
79	-347.90	-240.50	-280.20	-269.30	-280.20	-283.62	39.43
80	-253.30	-324.80	-417.90	-203.30	-417.90	-323.44	96.44
81	-355.40	-339.20	-281.10	-359.50	-281.10	-323.26	39.23
82	-216.10	-311.50	-264.30	-362.80	-264.30	-283.80	55.57
83	-220.40	-325.70	-315.60	-382.10	-315.60	-311.88	58.13
84	-266.00	-346.00	-279.30	-291.10	-279.30	-292.34	31.28
85	-1,037.90	-287.70	-277.50	-339.20	-277.50	-443.96	333.01
86	-288.60	-288.40	-372.80	-339.20	-372.80	-332.36	42.32
87	-220.30	-266.80	-332.00	-290.40	-332.00	-288.30	47.20
88	-206.10	-242.30	-348.70	-245.70	-348.70	-278.30	66.11
89	-265.30	-337.60	-281.10	-231.20	-281.10	-279.26	38.45
90	-277.90	-279.70	-344.60	-231.20	-344.60	-295.60	48.77
91	-288.10	-269.30	-292.00	-313.80	-292.00	-291.04	15.82
92	-277.00	-300.50	-279.30	-337.40	-279.30	-294.70	25.71
93	-249.20	-383.00	-206.70	-325.60	-206.70	-274.24	77.81
94	-209.50	-265.00	-266.60	-325.60	-266.60	-266.66	41.06
95	-227.30	-290.40	-358.20	-327.40	-358.20	-312.30	55.10
96	-356.10	-243.10	-302.00	-253.90	-302.00	-291.42	45.14
97	-243.20	-337.50	-291.10	-350.30	-291.10	-302.64	42.68
98	-281.70	-277.70	-300.40	-266.60	-300.40	-285.36	14.80
99	-273.60	-373.80	-292.00	-266.60	-292.00	-299.60	42.97
100	-240.60	-289.80	-323.40	-287.90	-323.40	-293.02	34.02

Table C.5: The realistic MBIE agent with 5 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-830.10	-731.70	-1,185.60	-960.90	-1,067.90	-955.24	181.26
2	-1,065.80	-891.60	-669.40	-429.90	-398.80	-691.10	289.28
3	-456.90	-409.70	-468.70	-480.50	-492.30	-461.62	31.88
4	-551.30	-433.30	-421.50	-456.90	-468.70	-466.34	51.03
5	-456.90	-421.50	-268.10	-480.50	-433.30	-412.06	83.61
6	-362.50	-492.30	-409.70	-504.10	-362.50	-426.22	68.60
7	-386.10	-433.30	-433.30	-338.90	-374.30	-393.18	40.53
8	-338.90	-362.50	-338.90	-433.30	-327.10	-360.14	42.87
9	-433.30	-409.70	-374.30	-421.50	-362.50	-400.26	30.54
10	-433.30	-327.10	-338.90	-456.90	-409.70	-393.18	57.57
11	-386.10	-539.50	-445.10	-397.90	-409.70	-435.66	62.10
12	-551.30	-445.10	-397.90	-350.70	-338.90	-416.78	86.15
13	-338.90	-268.10	-362.50	-350.70	-409.70	-345.98	51.16
14	-338.90	-468.70	-610.30	-421.50	-362.50	-440.38	107.70
15	-504.10	-433.30	-338.90	-350.70	-338.90	-393.18	73.50
16	-492.30	-456.90	-563.10	-338.90	-445.10	-459.26	81.50
17	-362.50	-445.10	-433.30	-456.90	-409.70	-421.50	37.31
18	-504.10	-433.30	-480.50	-445.10	-468.70	-466.34	28.17
19	-386.10	-433.30	-268.10	-468.70	-374.30	-386.10	76.02
20	-409.70	-480.50	-362.50	-350.70	-303.50	-381.38	67.06
21	-397.90	-433.30	-456.90	-527.70	-409.70	-445.10	51.44
22	-504.10	-480.50	-445.10	-433.30	-397.90	-452.18	41.38
23	-456.90	-480.50	-504.10	-421.50	-456.90	-463.98	30.77
24	-468.70	-456.90	-421.50	-386.10	-362.50	-419.14	45.24
25	-504.10	-268.10	-409.70	-362.50	-421.50	-393.18	86.55
26	-409.70	-492.30	-433.30	-397.90	-350.70	-416.78	51.84
27	-504.10	-445.10	-433.30	-504.10	-445.10	-466.34	34.80
28	-433.30	-421.50	-397.90	-409.70	-374.30	-407.34	22.70
29	-433.30	-504.10	-421.50	-386.10	-397.90	-428.58	46.16
30	-386.10	-386.10	-350.70	-386.10	-492.30	-400.26	53.69
31	-421.50	-409.70	-397.90	-374.30	-397.90	-400.26	17.50
32	-492.30	-480.50	-397.90	-492.30	-350.70	-442.74	64.85
33	-409.70	-374.30	-350.70	-433.30	-397.90	-393.18	31.88
34	-515.90	-527.70	-386.10	-421.50	-362.50	-442.74	75.28
35	-397.90	-315.30	-386.10	-480.50	-480.50	-412.06	70.01
36	-504.10	-515.90	-456.90	-480.50	-386.10	-468.70	51.44
37	-456.90	-397.90	-362.50	-456.90	-456.90	-426.22	43.84
38	-350.70	-350.70	-563.10	-327.10	-480.50	-414.42	102.73
39	-445.10	-480.50	-480.50	-409.70	-480.50	-459.26	31.66
40	-386.10	-456.90	-468.70	-421.50	-386.10	-423.86	38.60
41	-468.70	-374.30	-480.50	-433.30	-492.30	-449.82	47.64
42	-563.10	-397.90	-397.90	-397.90	-539.50	-459.26	84.43
43	-397.90	-350.70	-362.50	-456.90	-350.70	-383.74	45.24
44	-386.10	-374.30	-409.70	-291.70	-468.70	-386.10	64.09
45	-350.70	-397.90	-350.70	-350.70	-386.10	-367.22	23.00
46	-421.50	-504.10	-386.10	-362.50	-574.90	-449.82	88.15
47	-433.30	-315.30	-480.50	-386.10	-456.90	-414.42	65.49
48	-468.70	-374.30	-433.30	-315.30	-409.70	-400.26	58.64
49	-409.70	-386.10	-374.30	-456.90	-409.70	-407.34	31.66
50	-445.10	-386.10	-456.90	-445.10	-362.50	-419.14	42.05
51	-445.10	-504.10	-350.70	-327.10	-421.50	-409.70	71.78
52	-409.70	-338.90	-362.50	-338.90	-492.30	-388.46	64.85
53	-374.30	-492.30	-456.90	-504.10	-433.30	-452.18	51.84
54	-433.30	-374.30	-421.50	-433.30	-433.30	-419.14	25.58
55	-468.70	-515.90	-374.30	-539.50	-468.70	-473.42	63.33
56	-586.70	-338.90	-409.70	-421.50	-504.10	-452.18	95.35
57	-397.90	-397.90	-445.10	-445.10	-445.10	-426.22	25.85
58	-374.30	-433.30	-386.10	-504.10	-504.10	-440.38	62.22
59	-397.90	-421.50	-539.50	-421.50	-492.30	-454.54	59.24
60	-350.70	-303.50	-421.50	-397.90	-315.30	-357.78	51.16
61	-574.90	-338.90	-386.10	-362.50	-468.70	-426.22	96.44
62	-374.30	-445.10	-268.10	-374.30	-433.30	-379.02	70.11

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-374.30	-397.90	-362.50	-433.30	-445.10	-402.62	35.99
64	-386.10	-480.50	-386.10	-456.90	-386.10	-419.14	46.00
65	-362.50	-315.30	-445.10	-515.90	-386.10	-404.98	77.65
66	-468.70	-338.90	-315.30	-456.90	-421.50	-400.26	69.51
67	-421.50	-374.30	-421.50	-527.70	-409.70	-430.94	57.45
68	-421.50	-374.30	-433.30	-504.10	-504.10	-447.46	56.22
69	-397.90	-397.90	-409.70	-468.70	-445.10	-423.86	31.66
70	-468.70	-456.90	-362.50	-563.10	-468.70	-463.98	71.09
71	-350.70	-445.10	-456.90	-433.30	-409.70	-419.14	42.05
72	-492.30	-504.10	-386.10	-504.10	-421.50	-461.62	54.46
73	-445.10	-303.50	-409.70	-374.30	-409.70	-388.46	53.69
74	-338.90	-327.10	-456.90	-445.10	-374.30	-388.46	59.82
75	-433.30	-374.30	-433.30	-515.90	-445.10	-440.38	50.48
76	-445.10	-409.70	-374.30	-504.10	-456.90	-438.02	49.08
77	-338.90	-456.90	-492.30	-409.70	-374.30	-414.42	61.65
78	-515.90	-327.10	-504.10	-303.50	-468.70	-423.86	100.96
79	-362.50	-397.90	-433.30	-338.90	-362.50	-379.02	36.94
80	-409.70	-374.30	-409.70	-421.50	-445.10	-412.06	25.58
81	-350.70	-433.30	-468.70	-268.10	-386.10	-381.38	77.65
82	-480.50	-433.30	-433.30	-327.10	-468.70	-428.58	60.51
83	-409.70	-397.90	-386.10	-421.50	-433.30	-409.70	18.66
84	-397.90	-433.30	-327.10	-480.50	-468.70	-421.50	61.88
85	-397.90	-327.10	-421.50	-374.30	-421.50	-388.46	39.49
86	-409.70	-327.10	-445.10	-397.90	-338.90	-383.74	49.64
87	-492.30	-374.30	-421.50	-374.30	-374.30	-407.34	51.71
88	-374.30	-515.90	-480.50	-397.90	-492.30	-452.18	62.22
89	-409.70	-456.90	-433.30	-480.50	-433.30	-442.74	26.91
90	-445.10	-397.90	-374.30	-421.50	-480.50	-423.86	41.22
91	-409.70	-480.50	-468.70	-362.50	-374.30	-419.14	53.69
92	-468.70	-374.30	-374.30	-374.30	-362.50	-390.82	43.84
93	-397.90	-445.10	-409.70	-374.30	-386.10	-402.62	27.17
94	-362.50	-504.10	-350.70	-409.70	-350.70	-395.54	65.38
95	-456.90	-468.70	-445.10	-433.30	-468.70	-454.54	15.39
96	-374.30	-421.50	-338.90	-374.30	-386.10	-379.02	29.62
97	-327.10	-574.90	-433.30	-303.50	-515.90	-430.94	117.23
98	-468.70	-480.50	-445.10	-456.90	-456.90	-461.62	13.45
99	-374.30	-421.50	-445.10	-362.50	-433.30	-407.34	36.75
100	-433.30	-445.10	-386.10	-362.50	-421.50	-409.70	34.40

Table C.6: The realistic MBIE agent with 5 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-739.95	-401.95	-462.55	-452.65	-487.75	-508.97	132.84
2	-579.65	-804.45	-755.95	-977.85	-301.95	-683.97	256.36
3	-906.05	-819.35	-816.65	-588.75	-544.65	-735.09	158.64
4	-821.05	-1,032.25	-777.05	-924.15	-878.05	-886.51	98.74
5	-837.95	-833.15	-835.75	-921.05	-1,090.55	-903.69	110.83
6	-805.55	-1,118.35	-880.15	-556.35	-924.25	-856.93	203.96
7	-702.45	-998.35	-309.95	-1,315.65	-584.25	-782.13	387.25
8	-810.35	-670.55	-607.05	-1,642.05	-611.25	-868.25	440.31
9	-960.95	-871.05	-793.85	-702.35	-653.85	-796.41	124.32
10	-1,130.65	-521.65	-1,431.75	-621.05	-815.75	-904.17	375.48
11	-550.75	-638.05	-1,053.45	-761.55	-691.35	-739.03	191.89
12	-844.85	-700.75	-777.75	-500.15	-652.65	-695.23	131.43
13	-487.45	-884.55	-664.15	-584.55	-649.85	-654.11	146.49
14	-478.85	-1,028.85	-2,106.15	-619.15	-1,717.95	-1,190.19	702.79
15	-355.05	-712.95	-767.15	-881.55	-780.35	-699.41	201.92
16	-367.15	-969.95	-536.35	-616.45	-963.95	-690.77	267.71
17	-863.85	-749.65	-700.05	-638.35	-735.45	-737.47	82.69
18	-453.85	-635.45	-506.05	-515.25	-589.55	-540.03	72.03

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-437.55	-955.45	-669.45	-593.65	-828.85	-696.99	201.83
20	-403.15	-492.15	-547.35	-646.25	-573.95	-532.57	91.16
21	-286.95	-599.85	-648.75	-685.85	-288.95	-502.07	197.83
22	-301.25	-594.85	-630.45	-490.85	-554.55	-514.39	129.94
23	-297.85	-728.85	-783.35	-1,354.75	-333.95	-699.75	427.84
24	-624.95	-894.45	-678.95	-482.15	-416.05	-619.31	186.69
25	-312.75	-973.55	-518.95	-447.75	-403.75	-531.35	258.24
26	-299.75	-1,320.25	-784.25	-898.25	-229.55	-706.41	450.67
27	-371.25	-750.95	-735.65	-315.35	-396.65	-513.97	211.47
28	-391.55	-335.95	-683.25	-335.05	-397.45	-428.65	145.37
29	-333.25	-367.45	-965.75	-286.95	-321.65	-455.01	286.96
30	-383.15	-323.95	-896.75	-392.45	-339.35	-467.13	241.88
31	-300.55	-301.25	-1,700.65	-458.65	-431.95	-638.61	598.15
32	-361.25	-475.75	-1,039.65	-244.05	-254.05	-474.95	329.37
33	-335.95	-384.75	-659.55	-347.75	-367.45	-419.09	135.71
34	-302.25	-312.35	-979.25	-418.35	-344.35	-471.31	287.57
35	-336.75	-336.65	-1,060.25	-417.65	-432.95	-516.85	307.03
36	-456.55	-356.65	-431.15	-369.25	-267.65	-376.25	73.64
37	-347.65	-650.85	-323.05	-408.35	-349.35	-415.85	135.06
38	-357.85	-348.45	-344.45	-318.25	-362.95	-346.39	17.36
39	-361.75	-349.35	-276.05	-404.95	-365.65	-351.55	47.06
40	-344.45	-409.05	-490.25	-346.15	-359.65	-389.91	61.91
41	-346.95	-332.05	-337.55	-297.95	-380.95	-339.09	29.83
42	-325.75	-382.95	-288.45	-310.55	-359.45	-333.43	37.86
43	-337.85	-405.85	-387.45	-301.35	-335.05	-353.51	42.42
44	-263.15	-380.45	-311.25	-300.45	-373.85	-325.83	50.19
45	-404.25	-360.25	-273.05	-326.75	-473.45	-367.55	76.17
46	-298.65	-464.85	-323.95	-403.15	-446.05	-387.33	73.47
47	-417.75	-370.25	-419.05	-422.35	-336.65	-393.21	38.23
48	-277.75	-360.45	-416.95	-311.35	-338.45	-340.99	52.52
49	-342.75	-310.55	-346.85	-452.05	-430.15	-376.47	61.14
50	-441.25	-419.45	-736.15	-436.95	-334.85	-473.73	152.89
51	-393.95	-381.55	-361.15	-323.95	-360.25	-364.17	26.60
52	-333.45	-372.05	-441.25	-406.45	-338.85	-378.41	45.72
53	-393.25	-350.45	-323.25	-310.55	-251.55	-325.81	52.22
54	-431.15	-313.95	-360.25	-404.95	-396.55	-381.37	45.43
55	-359.65	-394.05	-266.55	-359.65	-382.05	-352.39	50.22
56	-291.35	-290.35	-349.45	-382.05	-420.45	-346.73	56.87
57	-355.45	-391.55	-276.85	-365.75	-358.05	-349.53	43.07
58	-361.35	-382.25	-335.25	-476.85	-337.55	-378.65	58.16
59	-462.45	-454.05	-332.45	-358.95	-395.85	-400.75	57.19
60	-310.55	-371.15	-337.65	-370.85	-347.65	-347.57	25.33
61	-357.85	-427.45	-413.45	-276.75	-374.75	-370.05	59.28
62	-405.95	-394.35	-313.15	-343.15	-335.25	-358.37	39.90
63	-419.45	-267.65	-357.75	-297.85	-299.85	-328.51	60.40
64	-297.05	-407.45	-358.75	-264.35	-358.45	-337.21	56.50
65	-346.85	-345.55	-372.05	-358.45	-410.15	-366.61	26.58
66	-370.35	-394.25	-345.45	-393.15	-428.85	-386.41	31.00
67	-467.55	-427.65	-394.75	-322.65	-431.95	-408.91	54.69
68	-360.75	-441.35	-338.45	-444.65	-389.45	-394.93	47.47
69	-393.25	-419.25	-324.35	-324.95	-290.45	-350.45	53.61
70	-357.55	-369.85	-478.35	-418.25	-322.15	-389.23	60.55
71	-432.05	-310.55	-300.35	-369.75	-348.45	-352.23	52.73
72	-437.05	-334.75	-345.75	-323.35	-457.85	-379.75	62.74
73	-372.05	-381.65	-385.65	-437.85	-442.25	-403.89	33.41
74	-359.85	-333.55	-392.25	-347.75	-367.35	-360.15	22.04
75	-405.85	-368.55	-323.95	-334.05	-375.05	-361.49	33.03
76	-360.45	-1,571.25	-301.55	-335.05	-361.15	-585.89	551.37
77	-336.65	-384.95	-251.35	-323.35	-280.35	-315.33	51.68
78	-425.45	-331.65	-411.45	-463.95	-431.95	-412.89	49.32
79	-357.75	-322.35	-524.75	-423.55	-325.75	-390.83	85.18
80	-287.95	-370.75	-313.05	-384.95	-345.95	-340.53	40.11
81	-360.45	-335.75	-313.95	-314.85	-303.05	-325.61	22.79
82	-394.15	-456.45	-418.75	-345.95	-336.95	-390.45	50.03
83	-383.15	-254.95	-373.85	-464.95	-347.75	-364.93	75.51

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-322.25	-308.95	-325.75	-385.65	-636.75	-395.87	137.86
85	-383.95	-453.35	-394.85	-349.65	-347.75	-385.91	43.02
86	-357.95	-408.35	-325.75	-336.85	-313.95	-348.57	37.14
87	-394.75	-396.55	-381.35	-440.55	-332.55	-389.15	38.71
88	-392.85	-301.25	-391.75	-317.55	-455.65	-371.81	62.83
89	-373.85	-393.85	-323.25	-379.75	-394.25	-372.99	29.19
90	-372.05	-431.25	-327.55	-251.85	-430.15	-362.57	75.60
91	-393.85	-385.65	-383.05	-444.05	-310.35	-383.39	47.77
92	-942.95	-345.05	-397.95	-359.55	-347.75	-478.65	260.41
93	-333.95	-312.15	-345.05	-410.15	-348.45	-349.95	36.52
94	-359.55	-309.75	-372.05	-384.15	-335.45	-352.19	29.80
95	-384.75	-314.15	-359.55	-312.15	-253.15	-324.75	50.52
96	-372.95	-311.55	-336.95	-439.45	-311.45	-354.47	53.77
97	-321.15	-394.05	-406.05	-339.35	-336.35	-359.39	37.99
98	-432.25	-379.25	-346.15	-335.05	-410.15	-380.57	41.24
99	-335.45	-324.15	-393.35	-430.45	-398.35	-376.35	44.99
100	-386.55	-374.75	-369.65	-325.75	-387.45	-368.83	25.26

Table C.7: The MBIE agent with 10 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-4,620.35	-5,204.85	-4,452.25	-3,927.95	-4,792.05	-4,599.49	468.14
2	-4,185.15	-4,727.85	-5,347.55	-5,499.95	-4,759.05	-4,903.91	529.31
3	-5,156.45	-4,977.45	-4,222.65	-4,333.65	-5,992.15	-4,936.47	713.70
4	-5,083.75	-4,048.05	-4,548.75	-4,452.85	-4,480.45	-4,522.77	369.94
5	-4,565.05	-4,710.05	-5,921.25	-5,028.65	-4,287.75	-4,902.55	629.01
6	-2,534.05	-3,529.55	-3,685.25	-5,132.75	-3,939.55	-3,764.23	932.50
7	-5,916.35	-4,963.15	-5,602.05	-3,086.45	-5,089.15	-4,931.43	1,101.09
8	-2,682.45	-3,248.55	-4,157.55	-4,497.45	-3,793.85	-3,675.97	722.91
9	-3,233.35	-5,448.65	-3,934.15	-4,091.55	-4,562.35	-4,254.01	820.47
10	-3,178.25	-3,013.35	-3,951.75	-4,054.05	-4,640.15	-3,767.51	669.61
11	-4,473.05	-2,921.65	-4,664.05	-3,143.65	-3,188.55	-3,678.19	821.82
12	-3,518.95	-3,664.95	-4,352.55	-2,391.25	-2,718.85	-3,329.31	782.35
13	-3,124.15	-2,584.15	-2,718.65	-4,286.95	-3,460.45	-3,234.87	681.90
14	-2,597.45	-2,819.35	-3,927.85	-2,121.35	-4,677.85	-3,228.77	1,046.56
15	-3,431.65	-3,765.05	-3,443.05	-2,280.85	-3,573.75	-3,298.87	584.74
16	-3,637.65	-3,456.25	-5,782.05	-4,538.25	-2,270.45	-3,936.93	1,309.95
17	-2,389.45	-2,343.05	-4,005.05	-2,210.65	-2,772.25	-2,744.09	735.18
18	-2,155.45	-2,987.95	-3,298.65	-2,963.95	-2,555.95	-2,792.39	443.16
19	-2,711.55	-4,280.75	-2,912.35	-3,512.55	-5,030.45	-3,689.53	966.87
20	-2,548.65	-2,395.25	-2,966.65	-2,581.45	-3,413.85	-2,781.17	411.57
21	-2,981.45	-2,175.75	-2,350.55	-4,389.55	-4,419.45	-3,263.35	1,084.02
22	-1,688.15	-3,323.45	-2,375.55	-3,247.85	-2,487.85	-2,624.57	677.22
23	-2,222.95	-3,193.75	-1,865.85	-3,250.15	-3,364.65	-2,779.47	685.57
24	-2,595.85	-3,010.15	-2,987.45	-2,928.15	-3,270.75	-2,958.47	241.58
25	-3,317.95	-3,536.15	-3,384.65	-2,095.85	-2,712.55	-3,009.43	599.20
26	-4,059.15	-3,553.25	-1,916.15	-2,205.15	-2,533.35	-2,853.41	914.43
27	-2,937.95	-4,235.05	-2,447.85	-2,419.25	-2,824.45	-2,972.91	741.36
28	-1,741.45	-3,623.25	-2,031.45	-2,066.35	-2,393.75	-2,371.25	737.07
29	-2,727.05	-2,596.55	-2,826.25	-2,577.15	-3,364.15	-2,818.23	321.59
30	-2,045.55	-2,838.85	-2,705.65	-2,333.05	-2,978.55	-2,580.33	383.49
31	-2,492.35	-2,268.75	-2,166.85	-2,344.15	-3,654.05	-2,585.23	609.16
32	-2,028.25	-2,218.95	-2,292.25	-3,108.75	-2,698.45	-2,469.33	433.00
33	-2,281.15	-3,480.35	-3,254.95	-2,126.95	-2,733.75	-2,775.43	590.22
34	-3,019.65	-2,426.15	-2,571.65	-2,976.05	-2,420.15	-2,682.73	294.39
35	-2,715.15	-2,595.05	-2,394.85	-2,121.65	-2,384.05	-2,442.15	227.07
36	-2,465.75	-2,301.45	-1,634.15	-3,295.45	-1,634.95	-2,266.35	688.91
37	-2,662.35	-2,493.55	-2,527.25	-2,651.95	-2,811.85	-2,629.39	126.27
38	-2,218.95	-2,902.85	-2,156.25	-2,242.65	-2,951.65	-2,494.47	396.71
39	-3,707.35	-2,182.05	-1,819.35	-2,242.75	-2,466.15	-2,483.53	722.50

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-1,429.05	-2,210.75	-2,456.35	-2,210.95	-2,493.55	-2,160.13	429.69
41	-2,013.45	-2,715.55	-2,658.85	-1,845.75	-2,064.75	-2,259.67	399.10
42	-2,096.25	-3,436.75	-3,174.25	-2,567.25	-1,986.95	-2,652.29	641.66
43	-2,737.05	-2,093.85	-2,257.65	-1,977.15	-2,610.35	-2,335.21	327.74
44	-2,990.05	-1,718.05	-1,592.45	-2,142.15	-2,336.55	-2,155.85	556.29
45	-1,962.45	-2,169.25	-2,401.35	-1,742.05	-3,880.45	-2,431.11	846.24
46	-1,466.35	-2,451.75	-2,990.55	-2,644.65	-2,074.55	-2,325.57	583.35
47	-1,876.85	-2,882.15	-2,354.75	-1,858.15	-2,759.05	-2,346.19	478.60
48	-3,087.55	-2,151.65	-2,037.15	-2,107.05	-2,718.95	-2,420.47	461.39
49	-1,936.05	-3,437.55	-1,929.35	-1,748.55	-3,237.35	-2,457.77	809.65
50	-2,942.75	-2,703.15	-3,062.75	-1,482.25	-3,281.65	-2,694.51	709.16
51	-2,429.05	-1,901.05	-2,216.55	-2,438.75	-2,454.95	-2,288.07	237.34
52	-1,662.65	-2,180.85	-2,274.25	-2,293.65	-1,804.95	-2,043.27	290.10
53	-2,334.15	-1,686.25	-2,326.75	-1,905.45	-2,622.15	-2,174.95	374.11
54	-2,045.75	-2,193.95	-1,984.25	-1,731.25	-1,715.45	-1,934.13	207.04
55	-1,717.55	-2,466.15	-1,700.45	-1,738.25	-2,486.95	-2,021.87	415.35
56	-1,770.75	-2,147.65	-2,215.05	-1,991.05	-2,780.65	-2,181.03	376.24
57	-1,946.45	-2,037.85	-2,029.85	-1,873.65	-2,231.55	-2,023.87	134.09
58	-2,020.95	-2,005.75	-1,924.35	-2,793.75	-2,274.25	-2,203.81	354.88
59	-2,096.45	-2,046.05	-1,736.65	-1,810.95	-1,987.85	-1,935.59	154.86
60	-1,944.85	-2,670.45	-1,872.35	-1,719.35	-2,019.75	-2,045.35	366.64
61	-1,965.95	-2,516.05	-2,564.85	-2,150.05	-2,303.55	-2,300.09	250.45
62	-1,896.25	-1,913.25	-2,828.45	-1,657.85	-2,181.05	-2,095.37	449.72
63	-3,244.75	-2,417.65	-2,457.85	-1,679.75	-1,942.55	-2,348.51	598.34
64	-1,770.05	-2,509.15	-2,402.75	-2,580.95	-2,024.75	-2,257.53	346.71
65	-2,038.65	-2,561.65	-2,313.75	-1,821.55	-1,540.35	-2,055.19	401.02
66	-1,829.65	-2,066.15	-2,202.15	-1,681.25	-2,842.45	-2,124.33	449.53
67	-2,100.55	-1,466.75	-2,066.75	-1,850.95	-2,252.75	-1,947.55	304.66
68	-1,850.85	-1,842.65	-2,986.65	-1,835.85	-2,703.55	-2,243.91	557.89
69	-1,588.45	-1,547.75	-1,617.75	-2,290.95	-1,820.05	-1,772.99	307.97
70	-1,841.05	-1,783.55	-1,684.95	-1,979.45	-3,171.55	-2,092.11	612.76
71	-2,643.05	-1,799.45	-1,841.85	-2,174.95	-2,115.35	-2,114.93	337.87
72	-1,780.95	-2,205.25	-1,462.15	-3,447.65	-2,348.25	-2,248.85	756.10
73	-2,611.15	-2,114.95	-2,919.15	-1,502.95	-2,194.15	-2,268.47	537.57
74	-2,487.35	-1,543.25	-1,821.65	-1,586.75	-2,133.15	-1,914.43	396.81
75	-1,874.45	-2,380.35	-3,130.65	-2,759.85	-1,992.65	-2,427.59	524.89
76	-1,500.45	-2,124.45	-2,442.85	-1,726.95	-2,025.15	-1,963.97	363.95
77	-1,776.85	-2,332.65	-3,092.65	-2,045.75	-1,774.45	-2,204.47	547.33
78	-2,016.55	-1,662.05	-2,093.75	-2,588.75	-2,195.45	-2,111.31	333.97
79	-2,140.25	-1,976.05	-1,729.35	-1,528.25	-2,022.05	-1,879.19	246.76
80	-2,414.15	-1,733.35	-2,179.25	-2,153.25	-2,704.55	-2,236.91	358.47
81	-2,096.75	-2,017.35	-2,002.15	-1,364.75	-2,621.65	-2,020.53	446.53
82	-2,517.35	-2,222.15	-1,938.95	-1,470.45	-1,763.75	-1,982.53	404.80
83	-2,175.45	-2,110.25	-1,950.75	-1,773.25	-1,851.05	-1,972.15	169.62
84	-1,887.05	-2,327.75	-2,775.95	-2,889.15	-1,935.65	-2,363.11	463.08
85	-1,566.05	-2,027.95	-1,911.75	-2,017.95	-1,481.15	-1,800.97	259.01
86	-1,479.35	-2,122.95	-2,040.45	-1,622.45	-1,805.05	-1,814.05	271.80
87	-1,496.55	-1,650.25	-1,907.75	-1,861.85	-1,766.55	-1,736.59	166.55
88	-2,031.75	-2,295.05	-2,205.35	-2,032.45	-2,135.85	-2,140.09	113.59
89	-1,895.15	-1,387.85	-2,414.85	-2,032.95	-2,261.95	-1,998.55	396.08
90	-3,362.85	-1,651.75	-1,946.95	-2,118.35	-1,760.95	-2,168.17	691.23
91	-2,112.65	-2,281.85	-2,096.95	-1,581.25	-1,519.85	-1,918.51	344.32
92	-1,699.95	-2,599.55	-1,763.35	-1,582.35	-2,413.55	-2,011.75	461.05
93	-1,980.65	-2,396.15	-2,018.35	-2,030.65	-1,665.35	-2,018.23	259.28
94	-1,643.15	-1,799.95	-1,968.95	-1,418.65	-2,038.05	-1,773.75	250.90
95	-2,234.05	-1,879.15	-1,828.65	-2,101.45	-1,762.55	-1,961.17	198.67
96	-1,922.05	-1,988.15	-1,876.95	-1,902.85	-1,669.35	-1,871.87	120.46
97	-2,439.35	-1,896.15	-2,177.75	-2,025.75	-1,998.45	-2,107.49	211.13
98	-1,795.35	-2,025.55	-2,030.65	-1,744.65	-1,454.65	-1,810.17	237.65
99	-1,448.35	-1,921.05	-1,987.05	-1,740.55	-2,093.15	-1,838.03	252.75
100	-1,684.95	-1,716.95	-1,821.65	-2,159.25	-1,955.65	-1,867.69	194.28

Table C.8: The MBIE agent with 3 reaches and 2 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-187.30	-1,025.20	-858.30	-365.90	-760.40	-639.42	350.21
2	-1,087.80	-564.00	-1,680.70	-1,004.20	-1,382.90	-1,143.92	419.59
3	-1,060.70	-1,061.00	-1,025.90	-639.90	-899.60	-937.42	179.06
4	-760.40	-1,095.10	-1,594.90	-1,175.20	-558.40	-1,036.80	399.78
5	-563.00	-461.70	-509.80	-1,209.60	-1,070.20	-762.86	349.54
6	-545.90	-815.40	-691.10	-997.20	-458.80	-701.68	214.26
7	-419.80	-612.60	-480.40	-845.30	-277.80	-527.18	214.71
8	-931.50	-900.60	-1,206.20	-502.00	-839.90	-876.04	251.94
9	-941.30	-275.80	-993.90	-317.90	-553.90	-616.56	338.04
10	-1,137.80	-535.70	-873.60	-514.20	-1,050.20	-822.30	287.74
11	-999.30	-482.50	-519.90	-594.10	-435.50	-606.26	227.25
12	-966.00	-261.70	-1,158.70	-308.60	-790.90	-697.18	398.34
13	-1,039.50	-499.80	-858.00	-560.40	-885.30	-768.60	229.46
14	-581.90	-727.70	-541.00	-740.60	-802.30	-678.70	111.63
15	-929.30	-794.20	-972.70	-548.20	-598.40	-768.56	190.84
16	-535.80	-621.10	-278.60	-800.20	-549.20	-556.98	187.93
17	-878.50	-812.40	-397.80	-506.70	-967.20	-712.52	246.88
18	-504.60	-314.40	-374.50	-686.90	-601.60	-496.40	154.44
19	-860.50	-318.40	-535.40	-254.30	-796.00	-552.92	273.03
20	-1,045.50	-629.20	-378.00	-469.90	-832.80	-671.08	271.32
21	-782.10	-575.10	-424.10	-397.20	-1,147.50	-665.20	309.98
22	-1,289.10	-373.70	-234.60	-641.50	-387.00	-585.18	419.99
23	-662.20	-619.00	-663.00	-520.10	-634.90	-619.84	58.81
24	-1,048.00	-680.00	-578.60	-621.80	-640.50	-713.78	190.36
25	-977.30	-482.10	-679.30	-593.00	-571.40	-660.62	190.42
26	-587.60	-502.40	-473.30	-841.00	-793.80	-639.62	168.47
27	-1,236.60	-946.60	-594.60	-461.30	-645.10	-776.84	312.38
28	-450.20	-368.90	-657.50	-766.10	-490.70	-546.68	161.61
29	-799.90	-567.50	-390.10	-529.80	-891.90	-635.84	205.42
30	-1,220.20	-440.40	-557.50	-1,046.00	-773.00	-807.42	326.21
31	-1,103.30	-492.20	-417.10	-734.20	-579.90	-665.34	271.78
32	-986.30	-353.30	-259.50	-754.60	-686.50	-608.04	298.78
33	-744.10	-306.40	-614.10	-739.40	-550.40	-590.88	179.33
34	-1,146.20	-566.60	-781.80	-546.30	-885.30	-785.24	247.59
35	-668.60	-449.60	-295.10	-344.40	-388.90	-429.32	145.37
36	-579.00	-468.90	-460.50	-469.70	-1,027.40	-601.10	243.28
37	-560.00	-649.50	-768.60	-536.60	-413.30	-585.60	132.59
38	-839.50	-255.50	-656.00	-855.60	-939.20	-709.16	273.85
39	-933.10	-317.30	-553.70	-502.60	-824.50	-626.24	249.73
40	-673.20	-367.80	-586.20	-366.00	-813.50	-561.34	195.15
41	-486.20	-328.40	-727.30	-378.50	-721.60	-528.40	187.84
42	-1,110.10	-441.90	-335.60	-721.20	-506.40	-623.04	306.55
43	-652.60	-404.50	-595.00	-534.90	-462.40	-529.88	99.50
44	-736.80	-390.60	-546.50	-608.80	-521.70	-560.88	126.47
45	-968.30	-1,004.00	-360.30	-546.70	-455.60	-666.98	298.99
46	-916.90	-509.50	-805.80	-589.00	-1,144.20	-793.08	255.36
47	-577.20	-474.60	-531.00	-420.70	-490.00	-498.70	59.05
48	-848.90	-516.00	-546.80	-404.50	-654.40	-594.12	167.96
49	-519.80	-348.40	-546.70	-508.80	-582.30	-501.20	90.00
50	-976.20	-770.50	-902.50	-789.70	-622.30	-812.24	135.42
51	-640.40	-469.50	-652.00	-567.80	-522.70	-570.48	77.49
52	-728.60	-458.30	-494.50	-752.10	-818.00	-650.30	162.60
53	-758.00	-915.80	-1,270.50	-414.10	-559.00	-783.48	332.48
54	-679.00	-798.40	-512.20	-541.50	-571.30	-620.48	117.75
55	-839.70	-600.40	-591.60	-384.90	-1,267.40	-736.80	337.51
56	-566.70	-377.50	-432.90	-755.30	-460.00	-518.48	149.19
57	-669.10	-259.10	-464.70	-325.20	-921.70	-527.96	270.35
58	-914.20	-599.30	-448.90	-314.10	-681.10	-591.52	228.86
59	-970.80	-377.50	-478.80	-191.40	-1,040.70	-611.84	374.89
60	-453.80	-721.10	-465.50	-434.90	-395.10	-494.08	129.69
61	-489.00	-540.60	-494.40	-242.00	-496.00	-452.40	119.43
62	-531.80	-267.20	-553.40	-745.40	-773.80	-574.32	203.49

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-418.80	-172.60	-385.70	-501.60	-672.40	-430.22	181.81
64	-1,004.00	-719.90	-494.70	-506.80	-679.40	-680.96	206.70
65	-581.80	-406.40	-504.80	-1,136.90	-696.80	-665.34	284.24
66	-569.20	-452.50	-493.70	-668.30	-489.30	-534.60	85.93
67	-818.70	-642.20	-895.00	-801.70	-505.20	-732.56	156.88
68	-727.10	-591.10	-579.20	-711.70	-448.20	-611.46	113.50
69	-543.10	-689.40	-268.00	-355.90	-517.00	-474.68	165.38
70	-866.90	-687.10	-755.80	-627.00	-748.20	-737.00	89.36
71	-902.00	-740.20	-520.60	-526.50	-729.40	-683.74	161.44
72	-613.60	-685.50	-686.10	-858.20	-535.50	-675.78	119.40
73	-622.80	-321.00	-486.40	-551.60	-558.70	-508.10	115.20
74	-486.80	-538.10	-425.30	-833.70	-418.40	-540.46	171.05
75	-691.90	-640.00	-700.80	-400.00	-524.40	-591.42	128.02
76	-691.90	-235.20	-558.10	-960.20	-469.90	-583.06	268.57
77	-1,033.10	-831.60	-388.50	-564.10	-938.40	-751.14	268.11
78	-756.30	-338.10	-612.50	-401.90	-357.50	-493.26	183.21
79	-714.80	-671.10	-493.30	-384.60	-499.50	-552.66	136.85
80	-546.20	-553.00	-662.10	-585.30	-605.30	-590.38	46.75
81	-453.20	-438.30	-530.20	-422.40	-763.80	-521.58	141.59
82	-550.50	-452.80	-660.10	-943.70	-789.10	-679.24	193.79
83	-698.10	-394.10	-871.80	-513.30	-513.90	-598.24	187.62
84	-807.60	-538.60	-299.90	-705.60	-687.80	-607.90	197.15
85	-593.00	-486.10	-373.10	-330.60	-689.20	-494.40	149.40
86	-842.40	-539.90	-560.60	-211.10	-432.80	-517.36	228.57
87	-595.20	-550.40	-452.10	-295.10	-458.80	-470.32	115.30
88	-669.20	-530.90	-382.70	-442.10	-615.20	-528.02	118.36
89	-631.60	-539.60	-566.60	-334.40	-697.20	-553.88	137.04
90	-709.40	-513.10	-923.50	-596.80	-458.40	-640.24	184.48
91	-596.20	-609.00	-537.10	-845.80	-579.70	-633.56	121.71
92	-726.30	-548.90	-758.20	-328.50	-631.30	-598.64	171.95
93	-729.30	-780.00	-642.20	-446.20	-206.20	-560.78	235.55
94	-708.30	-610.20	-727.10	-433.70	-697.10	-635.28	121.29
95	-344.40	-363.50	-700.10	-663.90	-603.40	-535.06	169.04
96	-714.30	-360.90	-692.70	-313.70	-677.30	-551.78	196.94
97	-549.10	-354.20	-516.20	-493.70	-537.60	-490.16	78.90
98	-638.20	-816.70	-597.40	-434.60	-925.60	-682.50	192.20
99	-502.10	-389.00	-538.20	-383.40	-820.40	-526.62	177.83
100	-462.30	-419.50	-395.30	-761.70	-585.50	-524.86	151.30

Table C.9: The MBIE agent with 3 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-396.95	-1,385.95	-565.95	-768.95	-529.25	-729.41	390.49
2	-663.15	-1,414.45	-1,148.45	-1,339.85	-786.75	-1,070.53	332.93
3	-670.85	-779.15	-943.95	-605.55	-504.55	-700.81	168.63
4	-571.65	-1,044.15	-778.55	-954.75	-659.85	-801.79	197.35
5	-526.95	-1,369.15	-731.75	-710.35	-854.85	-838.61	318.88
6	-775.95	-415.65	-1,538.95	-859.15	-846.45	-887.23	406.81
7	-838.35	-1,082.15	-528.75	-632.35	-488.85	-714.09	246.31
8	-903.25	-410.45	-1,606.25	-926.25	-550.95	-879.43	463.38
9	-886.25	-989.75	-627.95	-610.35	-534.95	-729.85	196.63
10	-1,447.85	-623.65	-519.35	-1,101.85	-1,504.15	-1,039.37	455.53
11	-722.45	-1,124.85	-731.55	-721.75	-449.55	-750.03	241.18
12	-1,329.25	-491.75	-800.55	-1,206.15	-389.95	-843.53	417.96
13	-495.75	-902.85	-288.35	-1,022.45	-457.45	-633.37	313.42
14	-848.05	-1,928.45	-516.15	-723.65	-1,261.25	-1,055.51	558.67
15	-1,067.45	-1,328.05	-857.35	-824.55	-743.75	-964.23	235.88
16	-835.15	-649.45	-1,017.25	-954.95	-439.15	-779.19	236.20
17	-867.65	-962.35	-693.15	-427.15	-1,200.15	-830.09	290.24
18	-1,418.75	-491.75	-1,455.95	-381.85	-489.95	-847.65	540.32

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-462.45	-510.15	-733.55	-616.25	-715.05	-607.49	120.47
20	-657.55	-799.15	-656.95	-592.65	-729.05	-687.07	79.09
21	-429.65	-445.95	-653.15	-1,226.55	-435.65	-638.19	342.00
22	-351.85	-366.75	-312.75	-1,060.85	-639.25	-546.29	315.43
23	-819.25	-522.15	-916.55	-690.05	-720.15	-733.63	147.98
24	-426.15	-567.95	-915.95	-486.65	-495.75	-578.49	195.24
25	-870.15	-377.45	-389.45	-528.95	-761.45	-585.49	221.87
26	-768.65	-647.15	-420.25	-680.15	-375.15	-578.27	171.46
27	-390.35	-424.75	-730.75	-493.75	-1,261.55	-660.23	361.44
28	-501.75	-469.35	-1,121.55	-969.45	-882.15	-788.85	290.05
29	-972.25	-521.15	-781.65	-659.85	-284.45	-643.87	260.44
30	-850.05	-456.55	-791.85	-1,087.25	-462.25	-729.59	270.34
31	-441.05	-612.55	-650.85	-831.65	-488.75	-604.97	153.23
32	-284.05	-435.85	-319.05	-859.35	-524.75	-484.61	230.25
33	-678.35	-614.95	-578.95	-1,121.95	-1,318.65	-862.57	335.78
34	-464.75	-742.25	-489.55	-617.85	-530.75	-569.03	112.97
35	-660.05	-403.25	-989.05	-1,205.05	-837.95	-819.07	306.77
36	-1,195.55	-353.05	-621.85	-636.35	-356.05	-632.57	343.39
37	-502.35	-773.75	-724.65	-464.65	-755.75	-644.23	148.37
38	-530.15	-655.35	-779.05	-632.05	-592.95	-637.91	92.05
39	-1,085.55	-406.75	-640.75	-688.25	-623.35	-688.93	246.76
40	-625.95	-578.65	-455.95	-734.35	-706.15	-620.21	110.85
41	-598.05	-777.85	-1,717.75	-649.95	-525.75	-853.87	491.62
42	-606.75	-905.95	-356.65	-1,108.15	-330.65	-661.63	340.82
43	-254.85	-1,023.95	-508.05	-1,200.75	-630.85	-723.69	384.93
44	-887.35	-649.15	-1,072.65	-1,179.75	-863.15	-930.41	204.87
45	-533.25	-726.45	-338.85	-509.85	-570.95	-535.87	138.79
46	-880.75	-839.25	-444.75	-682.95	-382.85	-646.11	225.59
47	-814.65	-894.55	-305.85	-905.15	-853.75	-754.79	253.51
48	-972.05	-597.45	-874.55	-719.25	-1,200.25	-872.71	232.63
49	-604.55	-1,122.95	-945.35	-1,092.85	-633.35	-879.81	247.64
50	-750.75	-459.95	-479.35	-623.55	-464.65	-555.65	128.39
51	-358.75	-264.15	-782.25	-811.75	-707.05	-584.79	254.63
52	-435.45	-618.85	-868.65	-539.65	-384.15	-569.35	190.49
53	-381.45	-808.05	-312.05	-416.15	-824.25	-548.39	247.35
54	-511.15	-633.75	-971.85	-546.85	-450.95	-622.91	205.99
55	-522.85	-1,113.15	-472.05	-748.55	-613.75	-694.07	256.74
56	-699.75	-593.55	-417.05	-771.55	-449.65	-586.31	153.75
57	-397.65	-606.25	-1,031.45	-477.45	-527.85	-608.13	248.51
58	-639.15	-649.95	-620.05	-695.45	-733.05	-667.53	45.94
59	-738.95	-471.55	-1,219.85	-529.35	-583.85	-708.71	302.57
60	-886.35	-681.45	-585.45	-756.25	-426.05	-667.11	173.92
61	-629.65	-902.55	-568.15	-718.65	-640.05	-691.81	129.39
62	-857.45	-531.25	-952.75	-739.15	-298.65	-675.85	263.08
63	-846.95	-422.15	-613.65	-911.25	-418.75	-642.55	231.02
64	-543.35	-747.55	-552.75	-402.45	-754.25	-600.07	150.03
65	-944.55	-543.05	-646.05	-854.85	-274.55	-652.61	265.10
66	-1,417.05	-393.25	-260.95	-1,415.15	-627.95	-822.87	557.26
67	-965.15	-389.35	-523.95	-741.85	-493.85	-622.83	230.37
68	-614.45	-376.35	-413.55	-432.35	-411.65	-449.67	94.31
69	-296.25	-750.45	-694.65	-868.25	-788.45	-679.61	223.41
70	-586.55	-261.35	-1,136.15	-563.95	-423.75	-594.35	329.59
71	-607.95	-553.15	-777.85	-753.05	-359.45	-610.29	169.30
72	-818.65	-261.65	-1,197.15	-325.75	-385.15	-597.67	400.07
73	-679.55	-331.15	-476.05	-313.45	-382.75	-436.59	149.81
74	-543.85	-625.65	-263.45	-626.55	-197.65	-451.43	205.73
75	-934.45	-726.15	-1,186.85	-854.25	-421.75	-824.69	281.14
76	-327.25	-453.35	-798.35	-456.95	-1,087.05	-624.59	312.21
77	-462.65	-459.85	-916.85	-642.85	-755.95	-647.63	196.03
78	-762.75	-431.15	-499.75	-1,393.55	-512.35	-719.91	397.04
79	-688.25	-577.45	-419.65	-576.55	-401.75	-532.73	120.47
80	-970.45	-1,043.55	-796.95	-893.85	-619.25	-864.81	164.97
81	-316.45	-699.65	-1,099.15	-508.05	-375.35	-599.73	315.60
82	-464.55	-380.65	-873.15	-1,125.75	-950.05	-758.83	321.67
83	-360.25	-796.85	-416.75	-614.55	-646.95	-567.07	178.05

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-374.75	-511.55	-1,039.55	-803.65	-313.85	-608.67	306.08
85	-431.35	-1,147.25	-1,376.85	-529.35	-585.65	-814.09	420.55
86	-233.25	-621.45	-501.95	-805.05	-688.85	-570.11	217.97
87	-1,083.45	-493.25	-1,019.95	-914.45	-688.85	-839.99	245.09
88	-560.55	-378.75	-576.85	-350.85	-238.95	-421.19	144.58
89	-517.95	-358.55	-390.45	-522.65	-425.85	-443.09	74.41
90	-1,125.55	-275.95	-717.35	-346.35	-623.05	-617.65	338.49
91	-1,074.15	-557.15	-1,223.35	-371.25	-718.25	-788.83	354.70
92	-614.65	-1,048.45	-650.05	-1,141.05	-161.15	-723.07	391.61
93	-835.05	-162.75	-362.25	-574.35	-419.95	-470.87	251.33
94	-385.25	-965.85	-407.35	-503.35	-771.65	-606.69	252.81
95	-201.75	-535.65	-634.05	-759.85	-439.95	-514.25	211.13
96	-574.75	-1,323.15	-731.55	-460.85	-317.65	-681.59	389.45
97	-428.25	-662.45	-1,206.25	-673.65	-655.45	-725.21	287.68
98	-281.25	-513.55	-435.35	-574.35	-683.45	-497.59	151.14
99	-793.15	-352.05	-371.05	-652.25	-432.05	-520.11	193.80
100	-588.35	-474.45	-866.15	-927.75	-432.35	-657.81	226.69

Table C.10: The MBIE agent with 4 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-3,290.80	-2,360.20	-2,236.40	-884.40	-2,099.50	-2,174.26	859.42
2	-2,807.00	-1,424.20	-3,586.70	-3,600.80	-1,199.30	-2,523.60	1,154.70
3	-2,464.00	-3,151.10	-617.70	-1,683.00	-460.40	-1,675.24	1,161.31
4	-2,564.00	-2,185.20	-660.40	-1,100.40	-2,987.10	-1,899.42	984.88
5	-1,549.00	-2,446.90	-1,740.10	-751.00	-451.00	-1,387.60	799.20
6	-1,560.70	-565.00	-3,458.00	-3,518.10	-1,591.20	-2,138.60	1,299.39
7	-2,846.70	-1,590.60	-309.10	-2,730.10	-1,948.80	-1,885.06	1,026.31
8	-581.10	-603.30	-1,204.80	-1,021.40	-1,411.80	-964.48	366.92
9	-1,406.40	-1,447.90	-1,709.00	-2,177.10	-925.50	-1,533.18	457.76
10	-758.80	-518.30	-1,594.40	-835.90	-669.20	-875.32	418.97
11	-1,910.10	-394.70	-964.30	-1,374.70	-2,873.80	-1,503.52	946.02
12	-1,591.10	-611.80	-515.20	-901.80	-1,121.70	-948.32	431.92
13	-1,122.30	-1,413.60	-868.10	-385.40	-1,478.40	-1,053.56	446.00
14	-370.80	-481.60	-1,627.70	-1,091.40	-953.70	-905.04	505.98
15	-624.40	-588.90	-770.10	-1,253.70	-456.00	-738.62	308.88
16	-677.40	-846.30	-1,827.20	-252.20	-796.00	-879.82	578.91
17	-702.00	-2,247.40	-1,355.90	-1,161.80	-2,924.10	-1,678.24	894.39
18	-953.00	-664.90	-475.50	-1,140.20	-776.90	-802.10	256.52
19	-585.50	-407.70	-431.50	-335.10	-1,420.60	-636.08	447.95
20	-404.90	-311.30	-916.40	-736.80	-732.60	-620.40	252.85
21	-491.70	-497.20	-397.10	-2,223.30	-979.90	-917.84	764.49
22	-396.00	-550.60	-787.20	-1,449.90	-701.10	-776.96	404.68
23	-566.60	-1,411.10	-1,169.70	-1,137.60	-434.10	-943.82	421.02
24	-540.20	-1,305.70	-1,541.60	-890.30	-1,117.30	-1,079.02	384.96
25	-825.00	-457.40	-806.30	-607.20	-1,173.70	-773.92	269.94
26	-600.20	-301.40	-392.20	-674.80	-1,114.50	-616.62	316.73
27	-854.30	-962.90	-375.40	-752.00	-963.10	-781.54	243.40
28	-628.90	-468.30	-554.00	-604.20	-1,093.60	-669.80	244.74
29	-637.60	-1,215.90	-637.40	-447.30	-350.60	-657.76	335.78
30	-647.00	-900.90	-701.00	-641.20	-786.90	-735.40	109.45
31	-465.30	-1,113.00	-1,070.40	-790.10	-455.20	-778.80	316.17
32	-242.80	-1,980.90	-1,080.80	-1,905.80	-1,032.70	-1,248.60	716.69
33	-387.40	-690.20	-622.90	-982.00	-447.80	-626.06	234.28
34	-393.00	-915.90	-1,059.60	-408.30	-897.20	-734.80	311.49
35	-500.00	-506.60	-963.40	-428.80	-454.10	-570.58	221.95
36	-642.60	-503.70	-1,856.70	-750.50	-678.70	-886.44	549.77
37	-536.50	-718.20	-2,340.40	-905.40	-1,460.20	-1,192.14	729.22
38	-346.80	-349.90	-729.00	-803.50	-946.20	-635.08	273.14
39	-668.20	-534.80	-737.90	-845.10	-479.20	-653.04	148.76

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-380.20	-367.20	-355.60	-571.30	-684.90	-471.84	148.45
41	-824.30	-1,665.50	-1,090.80	-839.80	-465.70	-977.22	444.56
42	-770.40	-695.20	-361.50	-1,135.70	-889.40	-770.44	283.05
43	-453.80	-470.20	-833.30	-766.20	-784.40	-661.58	183.93
44	-465.00	-820.80	-380.30	-1,156.30	-1,658.20	-896.12	525.85
45	-549.50	-798.10	-454.70	-311.40	-791.80	-581.10	212.84
46	-287.50	-1,289.80	-585.50	-559.40	-861.40	-716.72	379.33
47	-868.70	-510.30	-831.10	-442.60	-1,246.80	-779.90	322.09
48	-405.60	-899.20	-532.50	-381.80	-842.80	-612.38	243.75
49	-287.20	-351.90	-794.00	-426.20	-368.30	-445.52	201.00
50	-856.20	-633.90	-1,778.40	-552.60	-543.40	-872.90	521.65
51	-1,183.00	-367.40	-848.40	-597.70	-466.70	-692.64	328.08
52	-294.20	-424.70	-1,044.10	-404.80	-501.50	-533.86	294.70
53	-1,067.00	-308.50	-784.20	-780.70	-1,327.00	-853.48	379.77
54	-301.80	-428.30	-286.20	-638.80	-1,299.10	-590.84	420.33
55	-1,568.60	-581.70	-1,088.90	-433.10	-549.00	-844.26	476.89
56	-376.70	-317.30	-929.80	-1,158.40	-1,346.00	-825.64	461.60
57	-686.60	-1,217.80	-707.50	-486.50	-701.70	-760.02	272.00
58	-495.50	-407.50	-582.20	-413.60	-880.80	-555.92	195.05
59	-549.40	-554.70	-666.50	-1,161.40	-612.00	-708.80	257.47
60	-340.30	-526.60	-412.40	-1,426.30	-716.20	-684.36	438.38
61	-468.80	-552.60	-365.50	-427.00	-786.80	-520.14	163.85
62	-873.00	-554.70	-960.30	-446.40	-894.60	-745.80	229.40
63	-1,121.20	-584.10	-472.40	-1,361.30	-678.80	-843.56	379.92
64	-471.90	-450.00	-553.10	-629.50	-975.30	-615.96	213.05
65	-434.30	-346.60	-933.00	-871.20	-541.60	-625.34	262.82
66	-780.90	-433.10	-609.50	-447.40	-985.60	-651.30	234.25
67	-515.70	-642.00	-751.00	-957.20	-627.70	-698.72	166.81
68	-451.70	-345.80	-587.20	-759.00	-579.70	-544.68	155.86
69	-525.50	-757.90	-1,018.40	-627.20	-880.70	-761.94	196.18
70	-950.10	-586.70	-523.20	-636.90	-1,213.70	-782.12	291.89
71	-424.80	-1,461.50	-540.90	-743.90	-1,166.40	-867.50	435.92
72	-525.40	-546.00	-409.50	-1,065.90	-927.40	-694.84	284.63
73	-364.20	-440.40	-254.80	-579.90	-1,689.20	-665.70	584.25
74	-773.80	-329.80	-1,423.50	-866.50	-1,740.40	-1,026.80	557.25
75	-349.30	-369.30	-720.40	-798.60	-682.20	-583.96	209.45
76	-464.10	-473.80	-1,303.30	-803.40	-564.00	-721.72	352.71
77	-421.60	-711.70	-418.90	-1,174.10	-668.10	-678.88	308.31
78	-425.90	-600.90	-831.40	-355.80	-640.90	-570.98	187.72
79	-394.30	-550.00	-449.60	-801.30	-322.30	-503.50	186.08
80	-645.30	-548.60	-668.50	-731.00	-470.90	-612.86	102.94
81	-487.50	-811.10	-412.70	-273.80	-716.20	-540.26	220.43
82	-869.30	-411.60	-648.00	-1,009.50	-521.90	-692.06	245.83
83	-533.10	-313.40	-355.80	-778.00	-447.10	-485.48	184.21
84	-637.20	-622.70	-956.20	-704.20	-437.10	-671.48	187.50
85	-644.20	-661.00	-697.30	-297.60	-627.00	-585.42	162.98
86	-785.50	-803.90	-503.90	-465.60	-479.50	-607.68	171.40
87	-500.50	-364.70	-481.80	-553.40	-390.90	-458.26	78.55
88	-776.70	-453.80	-403.50	-820.00	-661.30	-623.06	187.56
89	-730.10	-456.20	-381.40	-972.30	-642.80	-636.56	234.07
90	-537.20	-820.20	-431.00	-1,386.80	-432.20	-721.48	404.48
91	-631.00	-585.00	-419.10	-352.50	-482.90	-494.10	114.90
92	-448.70	-942.60	-400.70	-382.40	-400.20	-514.92	240.35
93	-401.80	-746.00	-422.60	-738.60	-530.40	-567.88	166.56
94	-435.90	-188.80	-599.80	-490.00	-620.80	-467.06	173.34
95	-525.40	-257.50	-602.60	-558.50	-523.70	-493.54	135.79
96	-398.70	-686.80	-503.40	-496.70	-548.20	-526.76	104.79
97	-657.20	-288.40	-670.60	-458.80	-528.10	-520.62	157.28
98	-541.30	-658.30	-656.10	-840.50	-514.20	-642.08	128.78
99	-498.80	-844.50	-332.80	-526.40	-295.50	-499.60	217.46
100	-251.50	-249.90	-483.60	-306.10	-465.40	-351.30	114.90

Table C.11: The MBIE agent with 5 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-1,456.50	-1,444.20	-1,090.00	-1,741.20	-1,708.50	-1,488.08	261.75
2	-1,315.00	-2,157.30	-1,474.60	-2,104.00	-2,028.80	-1,815.94	391.24
3	-1,974.00	-2,604.00	-1,579.40	-1,571.40	-1,705.70	-1,886.90	432.65
4	-2,073.10	-1,863.20	-854.70	-1,192.20	-1,606.70	-1,517.98	495.30
5	-1,573.70	-1,811.20	-628.10	-1,565.20	-1,541.40	-1,423.92	458.12
6	-1,251.10	-1,482.30	-1,415.50	-1,547.50	-1,918.70	-1,523.02	247.14
7	-1,361.80	-1,932.00	-1,090.70	-1,954.00	-1,488.20	-1,565.34	373.55
8	-1,664.00	-1,976.70	-1,245.90	-1,493.10	-1,779.00	-1,631.74	278.31
9	-1,368.20	-1,181.40	-1,572.50	-1,484.90	-1,308.20	-1,383.04	152.27
10	-1,419.90	-2,091.60	-762.70	-1,261.00	-2,260.40	-1,559.12	616.01
11	-1,142.50	-1,446.70	-676.10	-1,003.80	-1,986.70	-1,251.16	495.72
12	-1,181.20	-2,113.90	-966.80	-1,686.50	-1,565.10	-1,502.70	447.71
13	-1,660.80	-1,508.70	-920.00	-1,497.90	-1,299.60	-1,377.40	286.07
14	-1,694.20	-1,119.00	-979.50	-2,327.40	-1,553.70	-1,534.76	532.74
15	-1,205.10	-1,538.10	-957.30	-966.60	-2,119.40	-1,357.30	487.16
16	-1,183.00	-2,031.90	-964.80	-1,068.40	-1,931.90	-1,436.00	505.52
17	-1,536.80	-1,176.70	-1,223.00	-777.30	-1,718.60	-1,286.48	362.19
18	-1,706.80	-1,660.30	-1,441.50	-838.80	-1,788.00	-1,487.08	384.44
19	-1,713.20	-2,035.60	-1,336.60	-1,591.90	-1,552.20	-1,645.90	256.79
20	-1,154.50	-1,681.30	-1,252.20	-1,253.00	-1,246.20	-1,317.44	207.62
21	-1,449.20	-1,530.30	-1,347.30	-1,722.10	-1,435.50	-1,496.88	141.65
22	-1,369.50	-1,556.40	-1,004.20	-1,663.90	-2,176.90	-1,554.18	429.07
23	-1,485.00	-1,087.60	-773.80	-1,375.20	-1,602.40	-1,264.80	334.34
24	-1,492.40	-1,303.10	-808.00	-737.10	-1,509.50	-1,170.02	372.61
25	-1,733.10	-1,045.50	-1,342.60	-1,738.20	-1,733.40	-1,518.56	314.31
26	-1,953.90	-1,329.20	-873.50	-1,115.00	-1,405.60	-1,335.44	402.94
27	-1,352.30	-1,367.00	-1,350.50	-808.40	-1,455.80	-1,266.80	259.91
28	-1,043.10	-1,166.40	-1,021.40	-1,387.40	-1,308.80	-1,185.42	160.89
29	-1,066.60	-1,264.90	-844.80	-1,161.20	-1,433.20	-1,154.14	219.97
30	-1,621.90	-1,366.50	-893.70	-1,381.50	-1,154.60	-1,283.64	273.66
31	-886.10	-1,444.80	-1,158.00	-870.60	-1,219.10	-1,115.72	241.65
32	-1,203.30	-1,543.40	-699.20	-1,487.00	-1,684.20	-1,323.42	390.32
33	-1,258.10	-961.40	-658.60	-898.70	-1,642.60	-1,083.88	378.38
34	-1,153.10	-1,403.30	-771.00	-1,015.30	-1,444.20	-1,157.38	279.39
35	-1,195.30	-1,695.70	-1,471.30	-1,100.20	-1,053.40	-1,303.18	272.75
36	-965.30	-1,380.20	-780.10	-673.30	-1,395.00	-1,038.78	335.17
37	-1,026.10	-1,489.00	-693.00	-884.80	-1,079.00	-1,034.38	294.81
38	-1,047.40	-1,605.70	-1,050.10	-1,393.00	-1,292.20	-1,277.68	237.66
39	-1,361.40	-1,322.70	-1,527.00	-1,248.30	-1,237.20	-1,339.32	116.94
40	-1,138.60	-1,339.60	-717.50	-916.40	-886.20	-999.66	242.01
41	-1,214.70	-1,069.00	-1,357.20	-1,009.80	-1,723.50	-1,274.84	284.74
42	-876.70	-908.20	-801.50	-1,180.70	-1,554.60	-1,064.34	309.26
43	-1,328.10	-1,345.60	-1,094.80	-910.30	-1,201.90	-1,176.14	180.11
44	-1,441.90	-1,015.00	-846.00	-1,166.80	-1,683.10	-1,230.56	334.62
45	-1,091.20	-1,281.30	-640.80	-1,074.00	-1,251.10	-1,067.68	255.98
46	-1,132.80	-1,416.80	-676.60	-950.60	-1,259.90	-1,087.34	286.27
47	-1,479.70	-792.40	-966.30	-800.60	-1,429.10	-1,093.62	337.05
48	-1,250.70	-939.30	-679.50	-1,202.40	-1,338.50	-1,082.08	269.78
49	-1,113.10	-797.80	-1,017.80	-1,066.60	-1,420.60	-1,083.18	224.02
50	-1,163.90	-1,159.50	-699.60	-528.90	-1,143.10	-939.00	302.63
51	-1,224.00	-1,087.50	-551.80	-678.60	-1,535.50	-1,015.48	402.36
52	-1,201.30	-822.40	-716.90	-1,689.10	-1,224.10	-1,130.76	384.64
53	-1,035.50	-870.70	-1,020.90	-1,562.10	-1,069.60	-1,111.76	263.04
54	-1,137.10	-1,177.20	-954.60	-1,180.50	-1,089.00	-1,107.68	93.25
55	-801.20	-1,074.50	-953.50	-803.70	-1,288.20	-984.22	204.68
56	-898.00	-1,144.60	-863.80	-1,313.70	-1,221.70	-1,088.36	198.99
57	-1,228.70	-905.60	-781.90	-1,059.90	-731.60	-941.54	204.52
58	-674.80	-1,093.60	-908.00	-1,083.20	-998.80	-951.68	171.92
59	-1,233.60	-1,241.10	-834.10	-1,179.00	-706.20	-1,038.80	250.53
60	-731.50	-936.70	-953.90	-490.20	-1,121.20	-846.70	242.54
61	-1,236.20	-727.10	-1,101.90	-1,128.60	-1,336.00	-1,105.96	231.27
62	-936.10	-1,264.90	-875.60	-1,037.20	-923.00	-1,007.36	155.54

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-938.40	-954.50	-529.80	-1,102.90	-881.30	-881.38	212.88
64	-1,043.70	-948.40	-587.80	-987.70	-942.30	-901.98	180.22
65	-1,156.00	-1,140.10	-660.40	-793.70	-1,427.90	-1,035.62	307.65
66	-870.40	-1,336.30	-873.70	-902.70	-942.40	-985.10	198.44
67	-767.20	-1,068.40	-957.80	-1,261.40	-1,738.60	-1,158.68	370.34
68	-842.80	-1,039.10	-745.60	-1,041.70	-824.70	-898.78	134.35
69	-851.30	-1,046.60	-620.50	-928.70	-856.90	-860.80	155.71
70	-1,069.60	-608.60	-788.20	-878.30	-879.10	-844.76	167.20
71	-1,013.80	-1,012.60	-563.30	-895.20	-760.80	-849.14	190.70
72	-1,061.00	-1,155.10	-579.70	-865.30	-738.50	-879.92	233.91
73	-1,073.40	-1,177.60	-471.70	-832.10	-957.00	-902.36	273.13
74	-917.80	-854.90	-438.20	-815.80	-1,068.00	-818.94	233.49
75	-939.70	-784.80	-671.50	-829.90	-1,021.60	-849.50	136.02
76	-1,030.20	-1,085.20	-1,192.80	-1,051.10	-1,100.50	-1,091.96	62.79
77	-836.60	-1,020.40	-636.80	-572.00	-1,135.70	-840.30	241.41
78	-815.70	-916.40	-1,309.30	-726.30	-971.40	-947.82	222.90
79	-1,115.20	-1,090.60	-824.00	-1,068.30	-1,043.10	-1,028.24	117.25
80	-968.70	-888.40	-600.30	-862.30	-1,332.10	-930.36	263.72
81	-1,009.10	-770.50	-729.00	-786.50	-1,087.40	-876.50	160.59
82	-739.00	-823.50	-523.00	-1,087.30	-847.80	-804.12	203.56
83	-1,289.20	-1,088.70	-844.00	-601.90	-932.20	-951.20	258.37
84	-1,265.90	-1,124.40	-974.40	-572.10	-1,114.90	-1,010.34	265.80
85	-863.90	-924.40	-901.80	-871.30	-769.00	-866.08	59.44
86	-1,342.60	-737.20	-562.00	-929.00	-1,066.40	-927.44	300.59
87	-789.70	-1,049.40	-1,171.30	-991.30	-714.20	-943.18	188.17
88	-659.30	-731.10	-1,048.50	-590.40	-1,087.40	-823.34	229.19
89	-988.20	-629.30	-734.10	-871.50	-1,231.10	-890.84	233.77
90	-997.60	-914.60	-733.90	-648.10	-953.50	-849.54	150.76
91	-1,206.30	-1,050.70	-695.40	-899.50	-1,067.40	-983.86	194.46
92	-1,021.90	-1,074.80	-567.60	-1,041.60	-784.20	-898.02	217.57
93	-864.80	-943.50	-653.90	-1,080.50	-884.00	-885.34	154.48
94	-1,054.90	-982.60	-680.00	-851.30	-1,066.00	-926.96	162.41
95	-691.90	-683.90	-408.10	-980.10	-1,046.50	-762.10	257.28
96	-921.00	-961.70	-849.50	-803.10	-1,033.90	-913.84	91.05
97	-1,193.40	-962.60	-799.20	-937.90	-876.10	-953.84	148.05
98	-735.80	-697.80	-875.30	-768.70	-988.80	-813.28	118.30
99	-830.40	-1,141.00	-631.10	-867.80	-828.20	-859.70	182.55
100	-800.90	-921.70	-876.50	-945.60	-663.60	-841.66	113.79

Table C.12: The MBIE agent with 5 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-3,731.95	-3,309.95	-3,104.05	-4,256.95	-3,308.75	-3,542.33	460.24
2	-4,445.35	-3,175.05	-4,306.45	-3,525.55	-4,557.25	-4,001.93	614.09
3	-4,377.65	-2,565.05	-4,461.75	-3,897.15	-4,315.15	-3,923.35	789.85
4	-3,902.55	-3,328.75	-1,490.05	-3,396.45	-4,111.95	-3,245.95	1,036.07
5	-4,260.25	-4,957.25	-1,797.15	-2,524.05	-2,923.95	-3,292.53	1,291.35
6	-2,781.95	-2,926.95	-3,563.75	-2,953.35	-3,419.55	-3,129.11	341.16
7	-3,812.85	-2,531.55	-4,178.75	-1,463.75	-3,495.85	-3,096.55	1,098.74
8	-3,509.65	-924.65	-927.65	-592.95	-3,758.95	-1,942.77	1,552.63
9	-2,806.75	-1,688.45	-3,316.45	-752.15	-988.65	-1,910.49	1,120.36
10	-2,870.65	-4,068.05	-1,292.75	-3,581.85	-1,218.55	-2,606.37	1,304.76
11	-1,154.55	-4,171.05	-3,674.55	-3,093.15	-4,165.25	-3,251.71	1,253.17
12	-873.35	-3,268.45	-959.35	-521.45	-3,589.75	-1,842.47	1,462.07
13	-4,485.65	-1,373.65	-2,987.55	-1,937.15	-2,410.25	-2,638.85	1,191.39
14	-2,800.65	-888.95	-2,130.35	-3,581.75	-2,601.45	-2,400.63	994.19
15	-4,431.05	-1,856.95	-3,334.95	-635.45	-1,851.15	-2,421.91	1,475.35
16	-2,880.95	-3,195.75	-769.65	-3,670.85	-2,477.55	-2,598.95	1,111.95
17	-604.45	-1,754.45	-1,538.45	-2,908.25	-3,844.75	-2,130.07	1,261.38
18	-1,284.05	-1,493.35	-454.85	-1,106.35	-2,399.65	-1,347.65	704.86

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-1,047.45	-1,174.35	-886.45	-1,905.15	-1,148.85	-1,232.45	392.69
20	-779.55	-420.65	-2,672.55	-1,258.55	-487.65	-1,123.79	926.61
21	-2,492.45	-4,214.45	-2,263.65	-1,818.25	-980.05	-2,353.77	1,189.89
22	-1,884.95	-1,137.65	-2,526.25	-1,152.85	-1,171.15	-1,574.57	619.18
23	-2,576.75	-1,988.55	-632.35	-2,128.15	-1,679.55	-1,801.07	728.78
24	-661.35	-1,417.35	-1,687.75	-3,327.45	-951.75	-1,609.13	1,039.95
25	-1,436.55	-893.45	-2,502.35	-1,909.65	-2,794.05	-1,907.21	772.97
26	-2,279.55	-1,811.05	-4,480.85	-610.15	-991.55	-2,034.63	1,517.49
27	-568.15	-3,127.35	-2,386.85	-903.05	-2,070.25	-1,811.13	1,060.74
28	-2,067.55	-1,597.85	-692.55	-2,239.65	-1,539.85	-1,627.49	602.30
29	-630.85	-1,512.25	-480.15	-477.05	-970.25	-814.11	438.89
30	-808.25	-1,259.95	-982.85	-880.95	-1,800.75	-1,146.55	403.93
31	-1,529.25	-2,085.55	-3,315.65	-1,202.85	-665.75	-1,759.81	1,010.83
32	-701.25	-898.85	-2,298.45	-636.95	-927.95	-1,092.69	685.47
33	-3,470.95	-1,099.15	-930.35	-3,378.25	-2,234.95	-2,222.73	1,206.87
34	-2,364.35	-2,627.25	-1,165.85	-1,201.15	-3,974.85	-2,266.69	1,162.39
35	-2,494.15	-1,471.95	-1,250.45	-1,831.25	-3,683.85	-2,146.33	980.00
36	-727.15	-2,100.95	-3,779.75	-625.75	-1,544.15	-1,755.55	1,284.08
37	-1,848.45	-2,317.15	-2,538.95	-2,405.05	-368.15	-1,895.55	892.53
38	-1,346.35	-2,244.35	-1,120.85	-3,250.85	-1,528.25	-1,898.13	865.33
39	-1,002.75	-980.95	-903.95	-855.25	-914.85	-931.55	59.94
40	-953.55	-1,205.25	-1,004.15	-1,179.75	-930.95	-1,054.73	128.84
41	-3,557.75	-1,416.15	-768.25	-1,948.05	-2,782.35	-2,094.51	1,101.68
42	-2,037.65	-502.15	-3,191.85	-1,374.05	-2,784.15	-1,977.97	1,080.48
43	-2,742.15	-1,569.45	-803.85	-1,547.65	-1,054.25	-1,543.47	745.64
44	-1,860.15	-3,051.15	-3,941.05	-3,499.65	-1,817.65	-2,833.93	961.40
45	-533.75	-516.55	-1,113.55	-1,113.25	-1,656.55	-986.73	476.19
46	-801.25	-1,506.55	-751.85	-820.05	-749.45	-925.83	326.08
47	-2,374.15	-1,068.55	-2,735.65	-1,389.85	-2,283.85	-1,970.41	706.62
48	-1,846.55	-1,736.35	-902.75	-2,041.55	-1,703.15	-1,646.07	436.04
49	-866.35	-678.45	-1,106.55	-740.45	-1,078.55	-894.07	193.68
50	-1,591.75	-2,789.95	-2,675.65	-1,433.45	-485.55	-1,795.27	955.53
51	-1,328.35	-900.35	-1,443.05	-1,417.65	-1,052.55	-1,228.39	240.01
52	-497.55	-1,021.25	-411.15	-2,345.35	-2,399.05	-1,334.87	975.48
53	-966.65	-3,273.75	-1,023.95	-2,422.05	-1,427.85	-1,822.85	998.94
54	-1,089.55	-968.45	-1,307.65	-589.25	-583.05	-907.59	317.62
55	-1,567.55	-4,265.55	-1,545.65	-4,031.95	-1,703.25	-2,622.79	1,396.76
56	-2,008.55	-722.05	-3,274.15	-1,147.15	-1,001.45	-1,630.67	1,036.72
57	-909.75	-898.45	-1,605.25	-1,231.85	-1,611.65	-1,251.39	352.37
58	-779.85	-683.95	-1,357.35	-818.35	-2,663.35	-1,260.57	827.09
59	-864.45	-3,230.75	-1,753.05	-615.55	-444.85	-1,381.73	1,149.90
60	-1,713.45	-418.05	-1,780.15	-1,317.65	-1,047.75	-1,255.41	555.27
61	-1,487.25	-830.35	-1,559.85	-433.05	-1,622.15	-1,186.53	527.77
62	-738.05	-713.15	-1,581.95	-776.45	-1,500.55	-1,062.03	438.99
63	-866.65	-1,065.65	-2,991.75	-2,991.75	-1,902.85	-1,963.73	1,015.81
64	-908.25	-1,192.55	-2,046.25	-787.75	-598.65	-1,106.69	567.75
65	-694.45	-959.45	-996.35	-1,048.95	-600.55	-859.95	199.31
66	-878.45	-446.55	-956.35	-1,882.95	-2,568.25	-1,346.51	860.47
67	-665.35	-1,115.95	-1,482.45	-575.15	-1,500.95	-1,067.97	437.76
68	-1,084.75	-1,263.25	-1,975.15	-646.55	-347.05	-1,063.35	624.22
69	-1,260.95	-754.45	-902.05	-1,089.35	-955.05	-992.37	192.16
70	-2,032.35	-1,116.75	-1,126.05	-1,370.85	-2,114.95	-1,552.19	487.68
71	-555.95	-684.35	-1,200.65	-1,064.95	-709.15	-843.01	275.07
72	-2,052.75	-2,189.25	-1,722.95	-1,467.55	-562.55	-1,599.01	644.43
73	-870.85	-613.85	-1,844.15	-1,022.35	-1,100.75	-1,090.39	460.48
74	-1,488.55	-1,873.15	-1,884.75	-1,016.35	-1,802.15	-1,612.99	370.40
75	-504.95	-567.75	-1,520.75	-745.45	-406.75	-749.13	448.65
76	-717.35	-1,016.35	-2,072.35	-888.85	-859.85	-1,110.95	547.85
77	-2,168.75	-698.55	-902.25	-1,030.85	-789.35	-1,117.95	600.45
78	-536.85	-1,419.95	-1,650.35	-903.05	-624.55	-1,026.95	490.04
79	-688.35	-1,276.15	-455.05	-630.05	-788.85	-767.69	309.02
80	-510.55	-738.05	-1,684.65	-978.45	-3,828.25	-1,547.99	1,348.60
81	-2,296.95	-872.45	-1,348.65	-800.75	-1,240.85	-1,311.93	598.15
82	-2,252.75	-581.35	-1,710.75	-1,396.05	-687.15	-1,325.61	702.57
83	-1,513.55	-2,052.35	-1,347.05	-1,754.05	-684.55	-1,470.31	513.26

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-1,667.95	-935.65	-2,056.75	-806.85	-2,300.15	-1,553.47	663.90
85	-884.35	-534.05	-1,431.15	-893.65	-542.95	-857.23	365.60
86	-1,411.45	-780.65	-1,070.35	-2,621.95	-1,234.05	-1,423.69	708.92
87	-756.15	-576.75	-2,536.05	-1,075.95	-607.45	-1,110.47	821.11
88	-505.15	-1,056.05	-2,088.65	-1,420.05	-1,236.35	-1,261.25	575.37
89	-506.35	-1,617.45	-3,171.25	-1,939.95	-743.55	-1,595.71	1,062.33
90	-746.45	-1,187.85	-698.95	-838.45	-1,633.05	-1,020.95	392.07
91	-1,969.85	-1,979.65	-511.25	-866.55	-1,358.65	-1,337.19	655.20
92	-2,046.55	-465.05	-681.95	-814.75	-2,588.25	-1,319.31	939.37
93	-846.15	-1,281.55	-950.05	-2,671.05	-1,302.85	-1,410.33	732.75
94	-521.15	-3,163.55	-1,242.45	-1,898.15	-1,425.75	-1,650.21	980.19
95	-1,015.85	-1,390.25	-2,100.45	-826.95	-911.55	-1,249.01	522.32
96	-1,794.95	-891.85	-1,062.55	-1,038.15	-2,334.45	-1,424.39	618.31
97	-1,981.35	-540.95	-2,764.35	-693.05	-1,028.45	-1,401.63	945.41
98	-710.55	-642.95	-619.35	-463.75	-2,213.05	-929.93	722.96
99	-659.65	-714.85	-1,355.45	-1,987.05	-697.45	-1,082.89	582.01
100	-3,224.75	-3,083.75	-1,486.55	-1,976.55	-641.45	-2,082.61	1,089.75

Table C.13: The realistic MBIE agent with 10 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-3,533.05	-4,538.65	-5,313.35	-5,661.55	-4,681.85	-3,954.58	818.30
2	-2,641.25	-5,732.25	-5,428.05	-5,982.95	-5,827.55	-4,268.34	1,401.72
3	-4,045.35	-4,624.75	-4,480.25	-3,918.75	-4,431.95	-3,583.01	302.31
4	-3,679.75	-5,193.55	-4,199.45	-4,976.25	-4,886.85	-3,821.98	629.01
5	-5,129.85	-4,180.05	-4,404.25	-4,373.65	-4,953.35	-3,839.36	409.62
6	-3,407.75	-6,074.25	-3,531.75	-4,128.65	-4,427.95	-3,594.06	1,069.93
7	-4,592.55	-3,962.15	-4,368.85	-4,081.65	-5,023.85	-3,670.34	424.30
8	-4,495.25	-4,558.75	-3,514.05	-5,062.45	-4,601.15	-3,703.94	567.40
9	-4,917.75	-2,880.15	-3,957.35	-4,621.35	-5,065.75	-3,572.23	894.98
10	-4,075.75	-3,004.85	-3,456.15	-4,078.45	-4,093.45	-3,116.44	493.26
11	-4,195.55	-4,104.85	-3,519.95	-5,580.05	-2,686.35	-3,345.96	1,060.71
12	-2,454.15	-2,962.45	-4,289.95	-4,470.05	-3,972.45	-3,022.84	878.46
13	-1,863.45	-2,909.75	-3,363.95	-2,578.15	-3,116.05	-2,303.06	581.08
14	-3,238.25	-3,899.95	-3,617.65	-4,082.05	-3,551.45	-3,062.56	326.17
15	-4,108.45	-4,569.25	-3,508.35	-3,555.05	-3,606.45	-3,222.09	459.69
16	-3,096.45	-2,995.85	-1,827.15	-3,226.45	-3,529.85	-2,443.29	651.13
17	-2,070.95	-3,371.05	-2,847.15	-2,616.95	-2,341.65	-2,205.12	497.56
18	-3,026.55	-3,542.35	-3,533.85	-2,281.65	-2,930.65	-2,549.51	519.81
19	-2,132.35	-2,118.05	-3,946.95	-3,105.35	-2,453.75	-2,289.57	779.03
20	-2,404.35	-3,100.75	-2,925.95	-3,030.35	-4,363.05	-2,634.08	723.41
21	-4,061.45	-2,536.35	-3,503.45	-2,849.05	-3,188.55	-2,686.31	590.35
22	-2,957.75	-3,043.85	-2,919.95	-2,971.15	-3,536.95	-2,567.94	256.10
23	-2,640.05	-3,568.05	-2,110.15	-3,470.15	-2,940.85	-2,451.04	602.90
24	-3,569.15	-1,861.25	-2,279.85	-3,060.05	-3,130.75	-2,312.84	692.54
25	-1,689.95	-2,315.55	-1,685.45	-2,419.55	-3,502.25	-1,931.29	742.84
26	-2,157.05	-1,775.25	-2,764.55	-3,036.75	-3,011.55	-2,119.86	559.16
27	-3,932.75	-2,514.25	-1,586.55	-4,042.25	-2,261.95	-2,385.12	1,077.86
28	-2,340.35	-3,142.55	-2,693.55	-2,448.55	-2,523.85	-2,186.81	314.18
29	-2,135.15	-2,695.05	-2,161.65	-2,549.85	-2,521.95	-2,005.77	250.26
30	-2,885.35	-2,672.75	-2,534.55	-2,279.35	-2,182.05	-2,087.34	286.67
31	-2,006.95	-2,505.55	-2,573.35	-2,009.75	-2,024.15	-1,848.12	289.08
32	-2,831.25	-3,655.55	-2,339.25	-3,598.95	-3,376.15	-2,628.19	562.73
33	-2,661.75	-2,058.35	-3,790.95	-1,921.45	-2,130.35	-2,088.31	767.84
34	-2,843.55	-2,050.15	-2,527.55	-2,614.85	-2,374.65	-2,062.79	295.15
35	-2,948.45	-2,180.55	-2,475.05	-2,504.65	-3,328.65	-2,233.73	451.31
36	-2,247.95	-2,344.65	-2,292.15	-2,655.85	-1,888.45	-1,898.84	273.75
37	-1,635.65	-1,726.85	-2,526.55	-2,283.05	-2,053.75	-1,698.14	373.30
38	-2,160.35	-2,078.75	-2,513.45	-2,654.25	-1,936.85	-1,884.27	302.69
39	-1,980.55	-2,756.15	-1,759.55	-2,175.75	-1,809.55	-1,740.42	403.34

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-2,620.25	-2,870.85	-3,103.65	-2,850.05	-2,941.65	-2,391.07	174.81
41	-2,401.95	-2,732.85	-2,879.65	-2,305.65	-2,398.85	-2,112.99	248.18
42	-1,841.15	-2,701.25	-1,805.15	-1,940.45	-3,011.55	-1,876.26	557.67
43	-2,016.35	-3,136.65	-1,885.85	-2,038.55	-2,160.15	-1,865.76	506.48
44	-1,840.25	-2,518.45	-1,764.25	-2,554.45	-2,356.25	-1,831.61	377.67
45	-2,500.95	-1,955.45	-2,072.35	-2,028.65	-1,280.35	-1,632.12	439.21
46	-3,003.05	-2,506.65	-2,449.45	-1,861.25	-1,784.85	-1,926.54	503.73
47	-2,041.45	-2,028.85	-2,023.95	-2,313.75	-2,059.95	-1,736.82	123.86
48	-2,390.05	-2,445.05	-2,323.35	-2,687.55	-2,273.95	-2,011.99	161.01
49	-3,114.25	-2,154.85	-2,831.15	-2,852.25	-2,283.05	-2,197.76	408.97
50	-2,141.65	-1,881.15	-2,313.35	-2,149.95	-1,864.05	-1,716.69	192.89
51	-1,735.85	-1,620.05	-2,365.65	-2,256.25	-2,291.25	-1,703.01	347.80
52	-1,865.95	-1,646.85	-2,391.55	-3,568.45	-2,672.15	-2,015.49	756.25
53	-1,587.15	-1,451.85	-2,696.65	-2,292.35	-1,888.35	-1,643.89	513.02
54	-2,359.55	-1,875.05	-2,060.25	-2,763.05	-2,888.95	-1,982.14	436.74
55	-2,292.15	-2,399.25	-1,821.75	-2,210.35	-2,285.15	-1,825.61	222.83
56	-1,820.35	-1,655.45	-2,642.45	-2,123.75	-2,466.35	-1,775.39	417.19
57	-1,567.75	-1,620.75	-1,650.95	-1,989.25	-2,252.85	-1,504.09	294.89
58	-2,098.95	-2,382.05	-2,151.95	-1,544.05	-1,549.95	-1,611.49	378.93
59	-1,718.35	-1,652.65	-1,830.85	-1,277.35	-1,744.45	-1,360.77	215.08
60	-1,719.35	-1,891.25	-1,679.75	-2,298.95	-1,987.25	-1,586.09	248.41
61	-2,765.65	-2,700.95	-1,735.65	-1,523.45	-1,173.85	-1,639.76	716.77
62	-3,115.95	-2,166.35	-1,481.85	-2,238.85	-2,081.25	-1,837.04	585.24
63	-1,923.75	-2,533.95	-1,357.05	-1,804.65	-2,036.85	-1,598.87	424.32
64	-1,421.45	-2,273.35	-1,740.45	-1,583.05	-2,359.55	-1,552.31	419.08
65	-2,332.95	-1,817.35	-1,776.95	-2,074.65	-1,244.55	-1,530.24	405.18
66	-2,000.15	-1,826.95	-1,516.25	-1,615.65	-2,833.05	-1,621.01	523.69
67	-2,826.85	-1,986.55	-1,485.25	-1,635.05	-1,479.45	-1,557.69	566.59
68	-2,017.55	-2,054.75	-1,810.85	-2,103.05	-1,730.15	-1,608.06	163.12
69	-1,791.45	-1,330.85	-1,583.05	-2,026.65	-2,070.75	-1,455.62	309.89
70	-1,667.65	-1,616.95	-1,480.85	-2,574.35	-1,923.25	-1,532.17	434.12
71	-1,596.55	-2,554.45	-2,247.85	-2,434.25	-2,215.65	-1,829.62	369.70
72	-1,997.45	-1,833.95	-2,388.25	-1,700.25	-1,651.45	-1,583.22	297.03
73	-1,467.15	-1,489.75	-1,737.35	-1,999.55	-2,508.45	-1,521.54	431.42
74	-2,285.15	-2,659.15	-1,733.75	-1,843.45	-1,754.65	-1,700.36	405.01
75	-1,461.65	-2,156.15	-1,764.75	-1,810.75	-2,028.65	-1,524.49	266.98
76	-2,532.05	-1,775.95	-2,330.55	-1,753.55	-1,231.55	-1,591.27	516.11
77	-2,017.25	-2,071.25	-2,175.55	-1,889.65	-1,301.55	-1,563.04	345.27
78	-1,404.65	-2,477.55	-2,787.05	-1,940.25	-1,709.45	-1,706.82	563.38
79	-1,306.35	-2,525.05	-1,405.25	-2,103.35	-1,934.25	-1,532.54	505.03
80	-1,987.95	-1,713.35	-1,401.95	-1,426.15	-3,716.95	-1,694.39	962.43
81	-1,764.35	-1,894.45	-1,927.35	-1,670.45	-1,933.75	-1,518.22	116.01
82	-1,823.55	-1,472.25	-1,603.45	-1,436.45	-1,877.35	-1,355.17	200.56
83	-1,255.85	-1,377.95	-1,550.35	-1,959.85	-1,468.35	-1,254.89	267.85
84	-1,537.85	-1,905.45	-1,940.55	-2,154.25	-2,222.65	-1,612.79	268.32
85	-1,817.75	-2,010.25	-1,415.55	-1,584.95	-1,403.55	-1,357.84	263.53
86	-1,454.55	-2,141.35	-1,710.45	-2,003.95	-1,466.75	-1,448.51	310.88
87	-2,562.35	-1,579.65	-2,149.75	-1,596.15	-1,279.25	-1,513.36	514.59
88	-1,795.15	-1,276.95	-2,185.65	-1,820.95	-1,835.95	-1,471.11	325.04
89	-1,893.75	-2,126.55	-1,955.85	-1,542.45	-1,962.85	-1,565.41	215.76
90	-1,655.25	-2,235.75	-1,853.55	-2,091.35	-2,679.35	-1,737.54	391.36
91	-1,349.15	-1,603.85	-1,523.65	-1,798.55	-2,035.85	-1,370.01	264.01
92	-2,231.95	-1,828.95	-1,610.55	-1,495.75	-1,793.75	-1,478.16	280.90
93	-972.95	-1,586.55	-1,623.55	-2,018.25	-1,365.95	-1,245.71	382.81
94	-1,619.55	-1,487.95	-1,425.35	-1,940.05	-1,414.05	-1,298.82	218.57
95	-1,435.95	-2,476.15	-1,361.85	-2,548.65	-1,373.75	-1,516.89	615.66
96	-1,530.45	-2,326.95	-1,152.35	-1,305.65	-1,818.75	-1,339.69	464.91
97	-1,712.65	-1,511.75	-1,821.55	-1,537.65	-2,110.05	-1,432.77	243.60
98	-1,743.05	-1,660.15	-1,282.65	-1,989.45	-1,297.95	-1,312.54	303.13
99	-1,851.65	-1,221.65	-1,295.75	-1,631.75	-1,724.25	-1,271.01	274.03
100	-1,633.55	-1,678.75	-2,023.45	-1,704.55	-2,025.25	-1,494.26	194.50

Table C.14: The realistic MBIE agent with 3 reaches and 2 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-602.70	-528.30	-630.20	-724.40	-1,271.30	-751.38	299.00
2	-967.30	-690.40	-702.60	-643.30	-650.80	-730.88	134.55
3	-742.20	-386.60	-935.40	-399.60	-532.60	-599.28	236.09
4	-309.80	-1,001.30	-509.20	-617.30	-911.10	-669.74	285.60
5	-395.20	-351.40	-238.60	-514.90	-492.90	-398.60	112.09
6	-703.40	-347.50	-343.00	-476.40	-206.70	-415.40	187.13
7	-241.50	-415.30	-495.90	-541.80	-321.70	-403.24	123.23
8	-231.60	-425.90	-325.70	-344.10	-313.20	-328.10	69.57
9	-391.30	-588.40	-306.50	-225.30	-356.00	-373.50	135.33
10	-345.00	-288.10	-208.00	-493.60	-206.90	-308.32	118.78
11	-201.20	-401.90	-361.20	-368.80	-306.50	-327.92	78.69
12	-355.10	-230.70	-366.30	-374.20	-208.30	-306.92	80.48
13	-336.20	-250.10	-470.10	-363.30	-230.80	-330.10	96.18
14	-288.20	-460.80	-305.80	-241.30	-251.70	-309.56	88.53
15	-227.70	-265.10	-337.60	-379.90	-422.50	-326.56	80.19
16	-219.80	-208.40	-325.90	-354.50	-330.30	-287.78	68.26
17	-239.50	-202.60	-283.00	-528.20	-230.30	-296.72	132.59
18	-264.20	-312.10	-370.80	-265.50	-385.80	-319.68	57.13
19	-239.80	-547.60	-166.00	-314.90	-379.40	-329.54	145.82
20	-359.70	-219.30	-292.00	-242.30	-275.50	-277.76	53.84
21	-281.40	-550.70	-221.00	-282.10	-394.40	-345.92	130.50
22	-357.20	-305.30	-309.00	-243.30	-344.80	-311.92	44.42
23	-243.00	-274.50	-425.30	-253.30	-431.90	-325.60	94.74
24	-265.80	-293.60	-368.50	-270.00	-261.50	-291.88	44.60
25	-298.10	-316.00	-275.90	-398.00	-185.80	-294.76	76.40
26	-458.50	-164.20	-153.10	-261.00	-382.00	-283.76	134.23
27	-279.50	-163.80	-332.80	-237.20	-450.60	-292.78	107.72
28	-274.60	-350.50	-331.80	-277.60	-373.30	-321.56	44.04
29	-371.70	-341.70	-288.60	-292.40	-230.30	-304.94	54.32
30	-465.10	-406.50	-339.30	-337.60	-313.10	-372.32	62.41
31	-404.40	-325.20	-314.30	-208.60	-351.70	-320.84	71.75
32	-303.30	-452.90	-236.60	-257.80	-219.30	-293.98	94.24
33	-252.20	-252.20	-275.70	-230.70	-340.20	-270.20	42.24
34	-371.70	-254.10	-292.30	-155.80	-306.90	-276.16	79.54
35	-209.40	-197.30	-251.10	-367.90	-264.30	-258.00	67.48
36	-213.20	-298.30	-206.80	-313.50	-300.20	-266.40	51.87
37	-226.60	-241.40	-242.70	-197.50	-142.30	-210.10	42.04
38	-298.20	-187.80	-516.10	-360.20	-511.10	-374.68	141.06
39	-178.10	-307.30	-208.30	-424.10	-174.30	-258.42	107.10
40	-264.20	-350.10	-152.00	-305.90	-320.00	-278.44	77.15
41	-234.00	-329.00	-490.40	-284.70	-374.20	-342.46	97.69
42	-350.80	-361.30	-297.00	-320.30	-349.80	-335.84	26.54
43	-272.70	-437.50	-291.80	-241.20	-232.00	-295.04	83.18
44	-181.50	-251.50	-254.50	-575.40	-429.10	-338.40	160.94
45	-372.10	-253.30	-307.50	-254.20	-240.40	-285.50	54.85
46	-230.00	-365.10	-252.30	-255.70	-359.70	-292.56	64.54
47	-408.40	-253.10	-256.50	-264.10	-352.20	-306.86	70.04
48	-242.70	-404.10	-253.90	-165.40	-341.80	-281.58	92.75
49	-403.30	-263.40	-419.00	-327.90	-231.60	-329.04	82.78
50	-410.60	-239.90	-219.20	-345.80	-347.20	-312.54	80.49
51	-371.30	-252.40	-311.60	-217.90	-337.90	-298.22	62.54
52	-233.00	-340.50	-318.30	-209.40	-262.80	-272.80	55.58
53	-317.70	-199.50	-383.80	-410.90	-325.60	-327.50	81.56
54	-227.50	-287.00	-233.20	-304.30	-316.40	-273.68	40.96
55	-286.20	-219.30	-263.30	-164.80	-320.50	-250.82	60.53
56	-255.70	-325.40	-353.90	-357.90	-221.00	-302.78	61.38
57	-275.40	-275.00	-185.80	-328.50	-206.80	-254.30	57.72
58	-316.40	-283.80	-251.80	-167.20	-394.10	-282.66	83.43
59	-376.50	-372.60	-176.50	-243.70	-186.60	-271.18	97.79
60	-166.20	-369.60	-176.50	-342.40	-200.30	-251.00	97.12
61	-273.20	-338.70	-253.30	-307.30	-287.20	-291.94	32.75
62	-356.80	-294.00	-220.00	-133.60	-185.80	-238.04	88.31

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-248.30	-416.10	-207.70	-245.80	-263.90	-276.36	80.81
64	-255.50	-340.50	-252.10	-377.20	-264.90	-298.04	57.20
65	-270.10	-251.90	-272.30	-307.50	-163.80	-253.12	53.84
66	-201.00	-320.70	-261.30	-201.90	-166.00	-230.18	61.10
67	-223.10	-221.50	-486.40	-322.20	-199.50	-290.54	119.34
68	-222.00	-300.30	-281.20	-199.30	-352.80	-271.12	61.64
69	-254.20	-294.90	-276.30	-330.00	-286.40	-288.36	27.81
70	-244.20	-255.00	-424.50	-322.20	-329.50	-315.08	72.22
71	-164.10	-348.60	-300.90	-276.80	-276.00	-273.28	67.78
72	-209.10	-373.90	-201.90	-298.80	-206.90	-258.12	76.23
73	-280.50	-230.80	-287.10	-331.80	-231.50	-272.34	42.47
74	-300.40	-296.80	-176.50	-256.90	-285.20	-263.16	51.36
75	-265.10	-309.30	-324.80	-210.60	-316.20	-285.20	47.64
76	-567.20	-259.20	-200.70	-366.10	-211.50	-320.94	152.42
77	-315.10	-252.60	-349.50	-257.10	-265.50	-287.96	42.52
78	-309.20	-151.90	-264.60	-289.30	-276.10	-258.22	61.70
79	-244.60	-317.40	-187.50	-274.70	-306.00	-266.04	52.29
80	-410.70	-220.10	-195.80	-243.20	-365.00	-286.96	95.00
81	-368.20	-339.80	-306.20	-201.50	-337.70	-310.68	64.86
82	-254.20	-287.40	-483.90	-387.70	-199.50	-322.54	113.28
83	-254.30	-426.70	-286.60	-309.00	-338.40	-323.00	65.63
84	-333.50	-273.20	-378.50	-363.90	-188.40	-307.50	77.88
85	-266.00	-273.80	-431.10	-343.60	-220.90	-307.08	82.06
86	-433.30	-339.70	-347.30	-144.50	-390.60	-331.08	110.84
87	-318.90	-217.90	-357.60	-187.30	-417.40	-299.82	96.05
88	-223.70	-510.10	-251.40	-306.40	-306.30	-319.58	112.35
89	-352.60	-419.70	-245.30	-308.60	-442.90	-353.82	80.77
90	-232.60	-229.10	-200.20	-270.30	-274.60	-241.36	31.08
91	-243.20	-297.90	-230.30	-280.80	-255.70	-261.58	27.57
92	-311.10	-285.30	-275.90	-342.70	-362.20	-315.44	36.81
93	-290.60	-185.80	-255.20	-300.80	-254.80	-257.44	45.07
94	-167.70	-277.20	-242.90	-371.80	-167.70	-245.46	85.25
95	-188.20	-262.40	-243.50	-384.70	-189.70	-253.70	80.20
96	-265.50	-276.80	-275.70	-278.40	-254.30	-270.14	10.20
97	-221.20	-424.80	-207.80	-213.60	-219.30	-257.34	93.76
98	-362.80	-223.70	-347.30	-212.80	-275.60	-284.44	68.90
99	-301.20	-365.60	-222.70	-319.20	-287.70	-299.28	51.96
100	-252.30	-277.20	-256.10	-366.70	-264.40	-283.34	47.57

Table C.15: The realistic MBIE agent with 3 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-947.15	-545.95	-815.35	-673.45	-1,294.75	-855.33	288.06
2	-1,268.85	-465.45	-684.45	-624.15	-743.15	-757.21	304.18
3	-578.95	-565.35	-755.85	-709.65	-631.75	-648.31	82.61
4	-981.15	-736.05	-785.75	-1,348.35	-762.35	-922.73	256.86
5	-948.05	-405.15	-484.45	-486.45	-709.65	-606.75	222.01
6	-408.35	-661.35	-259.25	-723.95	-682.15	-547.01	202.90
7	-528.55	-420.35	-404.35	-518.85	-592.55	-492.93	79.02
8	-306.15	-550.25	-492.35	-650.05	-548.15	-509.39	127.04
9	-1,100.55	-333.75	-1,112.45	-482.15	-339.85	-673.75	399.50
10	-535.85	-503.65	-588.05	-612.05	-1,069.15	-661.75	231.69
11	-435.25	-489.75	-427.35	-709.35	-694.45	-551.23	139.72
12	-553.25	-304.45	-257.75	-397.75	-397.15	-382.07	113.20
13	-685.45	-527.35	-285.65	-539.45	-438.05	-495.19	146.94
14	-507.85	-433.15	-170.85	-254.25	-314.15	-336.05	135.47
15	-750.15	-930.25	-1,050.95	-557.45	-399.35	-737.63	265.67
16	-642.15	-487.45	-499.95	-1,051.75	-457.25	-627.71	247.52
17	-1,003.05	-275.05	-586.65	-422.55	-195.65	-496.59	320.05
18	-668.35	-237.85	-779.85	-489.55	-216.45	-478.41	251.77

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-363.65	-517.55	-308.35	-280.85	-324.45	-358.97	93.58
20	-367.45	-364.45	-222.35	-643.75	-521.55	-423.91	162.18
21	-249.35	-463.45	-639.35	-303.75	-447.35	-420.65	152.79
22	-522.35	-292.75	-482.45	-735.75	-212.25	-449.11	205.66
23	-600.95	-602.35	-195.05	-318.65	-459.25	-435.25	178.36
24	-503.55	-594.25	-426.25	-360.95	-630.25	-503.05	112.46
25	-564.55	-821.95	-401.05	-425.75	-322.05	-507.07	196.56
26	-424.75	-462.75	-294.25	-602.85	-447.95	-446.51	109.96
27	-247.15	-637.35	-274.65	-450.15	-206.15	-363.09	179.37
28	-479.55	-876.95	-387.55	-227.95	-333.25	-461.05	249.69
29	-505.95	-640.55	-441.25	-324.85	-256.75	-433.87	151.02
30	-333.75	-382.55	-268.25	-526.55	-479.35	-398.09	105.29
31	-303.45	-433.65	-317.85	-414.15	-455.85	-384.99	69.63
32	-320.65	-375.85	-335.75	-466.95	-483.25	-396.49	74.76
33	-235.45	-416.75	-202.75	-748.05	-321.45	-384.89	219.34
34	-421.15	-379.95	-228.95	-195.75	-358.05	-316.77	98.68
35	-399.15	-423.95	-308.25	-184.45	-742.35	-411.63	207.35
36	-487.95	-353.25	-204.95	-417.15	-303.55	-353.37	108.04
37	-378.75	-348.25	-363.25	-302.65	-408.95	-360.37	39.32
38	-642.15	-270.95	-141.45	-343.35	-702.85	-420.15	242.40
39	-288.35	-575.85	-237.75	-450.85	-185.45	-347.65	161.73
40	-350.95	-247.35	-382.35	-531.05	-315.65	-365.47	105.29
41	-370.15	-328.65	-207.55	-319.85	-415.65	-328.37	77.51
42	-398.35	-376.75	-864.05	-247.65	-500.65	-477.49	234.10
43	-361.25	-574.25	-472.15	-212.35	-218.25	-367.65	158.18
44	-447.75	-440.45	-358.15	-235.75	-331.65	-362.75	87.15
45	-384.05	-333.65	-412.45	-389.55	-279.65	-359.87	53.28
46	-576.55	-249.35	-534.15	-293.95	-251.05	-381.01	160.85
47	-388.95	-535.65	-270.95	-298.55	-485.15	-395.85	114.71
48	-302.15	-265.15	-524.65	-223.45	-370.75	-337.23	117.92
49	-477.85	-488.25	-291.55	-313.45	-348.75	-383.97	92.79
50	-435.45	-463.05	-432.05	-185.75	-236.95	-350.65	129.01
51	-343.05	-242.65	-404.75	-173.55	-575.15	-347.83	155.22
52	-216.35	-237.55	-322.85	-365.85	-247.85	-278.09	63.39
53	-343.35	-271.85	-259.15	-295.85	-557.65	-345.57	122.84
54	-541.55	-331.55	-298.85	-277.95	-337.95	-357.57	105.71
55	-239.45	-336.15	-308.85	-310.75	-299.45	-298.93	35.92
56	-216.45	-416.75	-363.65	-251.45	-323.05	-314.27	81.44
57	-231.05	-321.25	-503.75	-397.45	-194.15	-329.53	125.56
58	-737.45	-450.45	-620.85	-244.45	-161.25	-442.89	243.51
59	-321.45	-184.75	-280.15	-342.75	-307.25	-287.27	61.66
60	-308.65	-277.05	-260.95	-335.45	-467.95	-330.01	82.29
61	-299.25	-373.45	-620.95	-248.75	-507.65	-410.01	153.00
62	-249.65	-529.75	-693.95	-280.55	-161.65	-383.11	221.08
63	-400.15	-495.75	-259.35	-512.35	-317.85	-397.09	109.86
64	-404.95	-193.95	-360.85	-173.65	-183.55	-263.39	110.44
65	-332.85	-331.65	-325.25	-173.05	-304.05	-293.37	68.25
66	-363.55	-172.55	-367.25	-251.55	-421.95	-315.37	101.05
67	-601.85	-270.65	-343.75	-443.45	-284.25	-388.79	137.20
68	-264.25	-264.75	-290.35	-280.05	-416.95	-303.27	64.49
69	-238.35	-349.15	-241.05	-247.95	-226.35	-260.57	50.13
70	-380.25	-230.15	-301.85	-757.45	-140.75	-362.09	238.05
71	-216.45	-185.75	-378.65	-344.15	-162.95	-257.59	97.41
72	-272.35	-427.35	-273.75	-318.55	-172.15	-292.83	92.34
73	-228.25	-311.55	-429.25	-265.55	-333.65	-313.65	76.44
74	-376.15	-453.85	-262.15	-395.65	-343.75	-366.31	70.67
75	-282.25	-489.85	-242.55	-340.35	-314.85	-333.97	94.52
76	-325.35	-436.45	-249.55	-239.15	-249.75	-300.05	83.72
77	-258.35	-392.65	-416.95	-390.45	-262.45	-344.17	77.19
78	-533.25	-259.85	-461.65	-334.35	-324.55	-382.73	111.48
79	-304.75	-226.25	-260.15	-425.15	-151.75	-273.61	101.50
80	-343.05	-390.15	-266.65	-467.65	-206.75	-334.85	102.25
81	-354.35	-322.75	-321.55	-451.35	-320.65	-354.13	56.17
82	-330.25	-288.95	-379.95	-420.55	-488.75	-381.69	77.82
83	-560.15	-313.95	-302.95	-259.95	-543.35	-396.07	143.66

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-465.45	-374.75	-357.75	-185.55	-257.75	-328.25	108.62
85	-240.75	-433.15	-452.05	-431.25	-311.75	-373.79	92.87
86	-457.25	-675.15	-462.65	-329.65	-301.45	-445.23	147.77
87	-395.65	-314.35	-265.15	-249.25	-297.65	-304.41	57.13
88	-293.95	-259.25	-485.35	-222.15	-440.55	-340.25	115.94
89	-268.95	-207.65	-318.65	-414.15	-260.15	-293.91	77.90
90	-354.45	-531.35	-341.55	-351.05	-333.95	-382.47	83.62
91	-459.45	-301.75	-403.55	-430.65	-175.75	-354.23	116.14
92	-289.95	-249.55	-216.95	-303.75	-397.55	-291.55	68.40
93	-371.35	-264.15	-294.15	-302.75	-152.45	-276.97	79.90
94	-608.85	-199.35	-248.55	-591.95	-181.65	-366.07	215.40
95	-240.05	-328.05	-267.65	-405.75	-249.45	-298.19	69.19
96	-248.45	-341.25	-335.45	-408.05	-340.35	-334.71	56.78
97	-257.45	-285.35	-340.95	-429.25	-185.45	-299.69	91.55
98	-362.25	-242.35	-507.15	-278.35	-440.65	-366.15	110.09
99	-227.05	-292.55	-239.35	-444.35	-490.05	-338.67	120.97
100	-270.25	-253.95	-257.55	-217.85	-358.85	-271.69	52.47

Table C.16: The realistic MBIE agent with 4 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-1,014.40	-1,124.20	-2,410.20	-487.50	-1,563.20	-1,319.90	719.80
2	-3,257.10	-3,359.30	-1,207.60	-1,799.60	-2,824.20	-2,489.56	945.65
3	-1,457.60	-1,747.80	-3,404.90	-1,967.80	-3,027.40	-2,321.10	847.44
4	-2,749.50	-487.90	-1,219.40	-3,220.10	-1,019.70	-1,739.32	1,179.77
5	-3,360.60	-648.20	-1,349.80	-3,171.90	-2,013.60	-2,108.82	1,163.58
6	-1,110.70	-666.10	-2,084.30	-981.50	-1,335.40	-1,235.60	532.62
7	-2,632.60	-1,366.60	-1,809.50	-3,285.40	-2,786.40	-2,376.10	774.77
8	-723.60	-1,031.70	-477.30	-667.30	-605.20	-701.02	206.29
9	-1,565.90	-1,185.30	-759.60	-1,036.10	-1,077.00	-1,124.78	292.30
10	-448.40	-2,169.50	-676.80	-2,250.10	-2,003.60	-1,509.68	872.86
11	-1,086.90	-586.20	-1,439.00	-988.60	-1,965.20	-1,213.18	518.66
12	-605.00	-484.00	-471.20	-559.20	-1,694.00	-762.68	523.51
13	-983.80	-915.00	-862.20	-2,378.10	-941.30	-1,216.08	651.08
14	-655.70	-374.40	-1,781.80	-1,133.90	-1,071.00	-1,003.36	534.80
15	-1,036.30	-1,361.60	-173.60	-564.20	-1,423.30	-911.80	535.07
16	-974.90	-804.50	-1,182.80	-495.10	-1,085.70	-908.60	270.54
17	-365.10	-522.90	-915.10	-334.40	-885.00	-604.50	279.32
18	-873.40	-872.40	-472.70	-508.40	-1,189.10	-783.20	296.97
19	-474.40	-1,931.40	-531.60	-343.20	-751.80	-806.48	645.93
20	-648.50	-642.20	-305.70	-250.00	-381.90	-445.66	188.22
21	-1,357.50	-1,667.30	-431.80	-1,321.70	-803.10	-1,116.28	492.60
22	-1,648.90	-749.10	-551.00	-500.80	-1,034.70	-896.90	469.81
23	-777.40	-497.90	-409.80	-669.90	-1,311.60	-733.32	353.72
24	-474.10	-1,693.40	-746.50	-687.90	-1,353.10	-991.00	510.71
25	-416.10	-1,482.90	-343.20	-469.20	-724.00	-687.08	467.40
26	-511.80	-1,408.50	-1,413.20	-911.70	-758.20	-1,000.68	400.69
27	-507.00	-669.60	-600.20	-773.80	-1,850.20	-880.16	550.97
28	-715.90	-605.80	-576.70	-594.70	-1,006.80	-699.98	179.96
29	-1,717.00	-875.90	-403.40	-383.20	-527.90	-781.48	559.01
30	-644.50	-626.00	-1,066.50	-605.20	-576.70	-703.78	204.33
31	-494.40	-404.60	-1,096.80	-735.80	-699.40	-686.20	268.01
32	-966.40	-1,763.90	-1,435.30	-434.30	-378.30	-995.64	608.40
33	-451.40	-490.60	-1,245.40	-896.60	-1,020.80	-820.96	343.34
34	-841.80	-609.20	-576.50	-630.70	-541.70	-639.98	117.74
35	-957.50	-668.30	-1,613.30	-780.10	-1,289.30	-1,061.70	387.62
36	-2,337.40	-491.50	-2,343.70	-536.60	-407.50	-1,223.34	1,020.92
37	-1,437.90	-2,812.70	-636.40	-704.90	-627.90	-1,243.96	940.45
38	-410.00	-594.70	-1,515.90	-628.30	-488.20	-727.42	449.20
39	-405.50	-1,180.80	-398.30	-376.10	-735.20	-619.18	347.28

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
40	-332.50	-1,426.00	-1,326.20	-539.30	-594.80	-843.76	496.94
41	-600.50	-768.40	-713.00	-823.90	-741.40	-729.44	82.88
42	-463.00	-1,062.00	-1,762.70	-345.50	-338.90	-794.42	618.07
43	-955.10	-976.80	-461.60	-523.20	-299.30	-643.20	305.87
44	-1,008.80	-516.70	-407.50	-487.90	-399.80	-564.14	253.64
45	-552.80	-1,223.80	-525.70	-1,146.50	-714.30	-832.62	330.92
46	-866.10	-715.10	-463.90	-397.80	-544.40	-597.46	191.39
47	-851.30	-521.50	-599.00	-1,226.10	-373.60	-714.30	334.38
48	-757.20	-679.20	-1,027.00	-454.60	-478.70	-679.34	233.29
49	-641.20	-669.10	-469.10	-1,018.60	-307.00	-621.00	265.74
50	-446.90	-753.20	-421.80	-345.90	-636.30	-520.82	168.23
51	-835.40	-540.00	-500.30	-495.20	-747.50	-623.68	157.23
52	-481.70	-793.90	-962.90	-621.10	-802.20	-732.36	185.08
53	-1,246.30	-707.60	-623.50	-522.80	-331.00	-686.24	343.14
54	-443.90	-1,938.90	-458.20	-820.60	-542.90	-840.90	632.24
55	-475.30	-934.90	-692.20	-671.70	-559.90	-666.80	173.67
56	-583.10	-332.40	-1,009.80	-258.50	-410.90	-518.94	299.73
57	-537.90	-530.70	-465.20	-565.00	-638.50	-547.46	62.72
58	-627.60	-831.20	-856.90	-380.80	-1,537.00	-846.70	430.68
59	-549.70	-482.70	-549.40	-692.50	-287.00	-512.26	147.38
60	-574.40	-1,115.40	-701.00	-342.90	-415.50	-629.84	305.03
61	-629.90	-885.70	-466.60	-542.50	-771.90	-659.32	169.98
62	-598.70	-528.80	-265.20	-592.60	-490.60	-495.18	136.24
63	-724.30	-857.70	-209.40	-515.40	-374.50	-536.26	260.83
64	-410.90	-600.90	-459.50	-865.30	-447.80	-556.88	186.92
65	-567.40	-506.50	-505.20	-351.00	-430.20	-472.06	83.34
66	-513.50	-481.90	-490.10	-527.50	-810.40	-564.68	138.56
67	-633.20	-621.60	-323.70	-483.60	-1,061.90	-624.80	274.65
68	-328.40	-1,015.90	-565.10	-791.00	-1,252.60	-790.60	363.55
69	-455.90	-599.50	-317.70	-341.50	-820.60	-507.04	207.81
70	-725.70	-594.70	-227.40	-406.10	-772.60	-545.30	227.53
71	-1,286.40	-374.80	-651.00	-567.60	-374.70	-650.90	375.27
72	-399.40	-1,886.40	-1,132.70	-562.80	-1,183.70	-1,033.00	588.11
73	-640.50	-374.00	-377.50	-520.00	-865.50	-555.50	205.68
74	-487.00	-413.30	-469.10	-827.20	-471.10	-533.54	166.52
75	-936.60	-825.70	-456.70	-399.70	-419.40	-607.62	253.58
76	-604.60	-555.80	-324.70	-690.70	-910.70	-617.30	212.74
77	-678.40	-570.80	-452.40	-1,474.20	-703.60	-775.88	402.83
78	-944.90	-385.30	-240.10	-542.50	-591.30	-540.82	264.79
79	-578.30	-443.00	-857.10	-400.90	-466.80	-549.22	184.19
80	-588.00	-1,843.00	-650.90	-968.70	-1,436.60	-1,097.44	535.35
81	-592.30	-391.10	-522.00	-450.40	-524.20	-496.00	77.18
82	-692.50	-565.20	-746.90	-338.60	-410.80	-550.80	175.62
83	-361.60	-481.90	-549.00	-365.80	-290.70	-409.80	103.71
84	-353.60	-491.10	-986.00	-497.20	-674.90	-600.56	243.80
85	-392.40	-621.20	-728.50	-324.80	-567.40	-526.86	165.91
86	-932.80	-552.80	-278.30	-475.30	-630.60	-573.96	239.57
87	-747.50	-1,134.00	-278.00	-305.20	-776.90	-648.32	359.56
88	-430.20	-745.70	-530.00	-621.90	-388.30	-543.22	145.03
89	-1,819.40	-627.80	-488.40	-544.10	-886.90	-873.32	550.47
90	-694.90	-569.80	-259.40	-478.70	-1,517.70	-704.10	481.83
91	-472.60	-479.60	-1,533.40	-284.40	-518.10	-657.62	497.91
92	-476.00	-808.50	-347.30	-401.10	-334.60	-473.50	195.38
93	-548.90	-504.40	-348.80	-698.30	-1,376.40	-695.36	400.58
94	-411.20	-690.30	-223.40	-424.20	-306.70	-411.16	176.25
95	-955.90	-372.40	-423.70	-644.00	-1,349.80	-749.16	406.95
96	-259.00	-638.60	-322.90	-557.80	-470.40	-449.74	158.31
97	-809.80	-1,051.00	-601.40	-511.20	-300.50	-654.78	287.23
98	-488.50	-425.90	-494.20	-401.20	-392.40	-440.44	48.11
99	-768.30	-466.30	-258.80	-347.90	-336.10	-435.48	200.30
100	-445.50	-361.10	-448.70	-959.70	-447.80	-532.56	241.68

Table C.17: The realistic MBIE agent with 5 reaches and 1 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-1,674.80	-1,210.30	-1,808.80	-1,373.90	-1,356.30	-2,727.70	-1,691.97
2	-867.80	-517.10	-1,108.50	-1,170.60	-956.00	-957.10	-929.52
3	-1,304.00	-902.30	-762.00	-938.30	-901.10	-1,873.70	-1,113.57
4	-761.80	-737.80	-704.70	-459.40	-952.60	-831.40	-741.28
5	-1,211.00	-659.10	-826.90	-851.60	-787.80	-832.70	-861.52
6	-698.40	-453.50	-786.60	-737.00	-418.50	-724.90	-636.48
7	-804.70	-608.90	-802.30	-578.40	-610.50	-922.60	-721.23
8	-761.30	-558.20	-654.70	-630.90	-1,017.20	-688.00	-718.38
9	-804.30	-720.00	-994.30	-900.60	-746.10	-468.20	-772.25
10	-665.30	-843.90	-643.90	-526.40	-516.70	-536.90	-622.18
11	-781.30	-804.90	-625.70	-775.70	-332.20	-956.10	-712.65
12	-567.90	-610.30	-494.70	-668.20	-512.00	-571.40	-570.75
13	-487.00	-648.20	-722.70	-618.50	-716.90	-698.90	-648.70
14	-752.90	-519.80	-742.80	-712.20	-840.10	-635.50	-700.55
15	-690.40	-598.20	-650.50	-633.90	-621.70	-600.10	-632.47
16	-580.40	-635.20	-725.20	-612.30	-596.70	-529.20	-613.17
17	-694.70	-659.70	-569.00	-619.50	-633.50	-660.80	-639.53
18	-669.20	-532.60	-560.30	-571.20	-511.40	-569.10	-568.97
19	-527.70	-557.50	-619.70	-567.90	-547.10	-521.60	-556.92
20	-487.80	-543.70	-419.00	-660.40	-575.40	-471.90	-526.37
21	-701.70	-636.70	-760.50	-526.50	-634.50	-650.20	-651.68
22	-573.80	-501.90	-596.60	-447.80	-645.20	-703.60	-578.15
23	-624.20	-663.60	-564.70	-425.00	-443.00	-308.70	-504.87
24	-783.30	-406.30	-624.70	-556.10	-434.30	-509.60	-552.38
25	-628.70	-430.80	-603.70	-777.10	-764.60	-535.10	-623.33
26	-633.30	-469.60	-584.10	-552.70	-621.80	-538.40	-566.65
27	-512.30	-445.40	-486.10	-643.90	-468.20	-552.20	-518.02
28	-533.40	-624.90	-459.90	-528.90	-753.20	-728.30	-604.77
29	-660.90	-471.10	-640.70	-650.10	-376.10	-485.40	-547.38
30	-729.70	-438.20	-498.50	-480.90	-560.00	-563.20	-545.08
31	-585.80	-315.10	-546.80	-521.60	-606.70	-764.80	-556.80
32	-551.70	-426.60	-475.80	-707.20	-422.60	-645.00	-538.15
33	-735.10	-534.10	-454.10	-642.20	-643.50	-668.20	-612.87
34	-693.60	-492.30	-701.40	-580.30	-464.80	-513.10	-574.25
35	-561.20	-347.10	-488.20	-467.20	-582.10	-464.90	-485.12
36	-473.20	-758.10	-599.20	-633.60	-728.80	-518.40	-618.55
37	-398.30	-495.50	-550.60	-677.70	-600.10	-586.30	-551.42
38	-590.50	-436.90	-520.60	-514.30	-515.90	-656.70	-539.15
39	-574.60	-603.30	-533.90	-579.60	-423.20	-557.10	-545.28
40	-489.50	-498.90	-513.10	-410.00	-636.30	-762.50	-551.72
41	-555.30	-395.10	-577.20	-479.10	-610.80	-452.40	-511.65
42	-322.10	-370.90	-511.50	-491.90	-660.60	-615.20	-495.37
43	-421.10	-699.30	-536.50	-356.40	-377.80	-608.30	-499.90
44	-718.80	-348.70	-621.00	-446.90	-392.80	-521.60	-508.30
45	-641.50	-411.60	-497.90	-313.50	-553.50	-657.10	-512.52
46	-470.90	-519.40	-643.70	-686.10	-732.60	-554.40	-601.18
47	-496.20	-417.20	-583.90	-499.20	-614.20	-588.20	-533.15
48	-500.40	-495.90	-475.10	-563.20	-560.20	-555.30	-525.02
49	-554.60	-549.70	-545.00	-420.90	-515.50	-734.40	-553.35
50	-580.50	-460.10	-535.90	-577.00	-489.10	-519.00	-526.93
51	-611.60	-413.90	-684.70	-593.50	-508.80	-673.40	-580.98
52	-527.30	-492.30	-598.40	-644.30	-585.00	-566.80	-569.02
53	-620.50	-505.80	-509.80	-579.60	-443.20	-448.60	-517.92
54	-475.90	-516.80	-563.10	-502.00	-517.80	-613.50	-531.52
55	-623.60	-484.80	-513.80	-544.40	-630.70	-673.60	-578.48
56	-341.70	-443.30	-590.00	-639.50	-458.70	-528.80	-500.33
57	-605.00	-572.80	-534.40	-501.70	-639.90	-466.60	-553.40
58	-617.30	-408.70	-476.90	-568.10	-527.40	-440.40	-506.47
59	-599.80	-675.50	-388.10	-411.70	-524.30	-388.80	-498.03
60	-466.60	-565.70	-436.10	-388.70	-647.00	-545.20	-508.22
61	-515.50	-527.90	-684.30	-520.30	-492.50	-636.30	-562.80
62	-546.10	-574.20	-536.40	-626.00	-542.00	-551.90	-562.77

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
63	-586.90	-340.30	-598.20	-459.70	-568.30	-448.60	-500.33
64	-465.80	-785.50	-442.30	-460.40	-470.20	-582.20	-534.40
65	-716.90	-463.40	-515.50	-630.70	-416.20	-366.00	-518.12
66	-520.70	-440.00	-520.90	-572.90	-394.40	-661.30	-518.37
67	-482.20	-489.60	-650.30	-640.50	-465.10	-610.20	-556.32
68	-427.60	-548.80	-613.50	-591.50	-507.60	-750.40	-573.23
69	-533.20	-612.00	-438.50	-369.90	-391.20	-492.70	-472.92
70	-626.60	-557.90	-399.10	-505.30	-597.00	-445.40	-521.88
71	-612.80	-431.40	-647.10	-502.70	-462.90	-502.10	-526.50
72	-630.30	-655.80	-641.30	-652.90	-607.20	-567.10	-625.77
73	-717.30	-496.50	-388.00	-413.30	-594.10	-619.40	-538.10
74	-393.70	-580.90	-486.00	-617.70	-483.80	-388.80	-491.82
75	-607.30	-543.60	-513.00	-504.50	-462.90	-590.30	-536.93
76	-474.30	-421.30	-710.10	-554.30	-361.30	-389.70	-485.17
77	-556.00	-354.30	-491.80	-495.90	-485.60	-623.50	-501.18
78	-498.70	-488.90	-636.50	-622.50	-507.60	-627.70	-563.65
79	-647.20	-542.90	-503.50	-528.70	-532.10	-470.70	-537.52
80	-479.20	-428.10	-610.00	-574.60	-622.30	-580.50	-549.12
81	-523.90	-246.70	-544.30	-518.50	-563.50	-458.10	-475.83
82	-425.90	-448.20	-508.10	-448.70	-472.00	-609.30	-485.37
83	-745.10	-606.50	-657.60	-676.80	-559.80	-554.40	-633.37
84	-498.80	-586.20	-533.40	-579.50	-602.00	-715.40	-585.88
85	-528.50	-609.50	-547.30	-483.70	-505.00	-591.50	-544.25
86	-447.10	-452.50	-564.20	-583.10	-555.60	-438.50	-506.83
87	-555.00	-528.40	-517.80	-469.80	-272.70	-929.50	-545.53
88	-561.80	-510.90	-435.20	-511.20	-512.50	-484.20	-502.63
89	-600.60	-485.70	-600.10	-582.00	-478.60	-570.40	-552.90
90	-619.60	-397.50	-564.80	-368.50	-598.00	-545.30	-515.62
91	-513.00	-650.30	-571.90	-395.30	-505.80	-469.80	-517.68
92	-514.50	-568.20	-355.10	-589.70	-490.50	-524.90	-507.15
93	-474.80	-444.00	-547.70	-482.20	-359.10	-510.60	-469.73
94	-412.50	-442.40	-422.70	-661.10	-498.90	-612.90	-508.42
95	-560.20	-436.70	-375.20	-574.50	-542.80	-576.40	-510.97
96	-389.70	-527.20	-376.90	-708.40	-486.50	-495.10	-497.30
97	-609.40	-359.90	-529.70	-618.10	-427.50	-616.60	-526.87
98	-559.40	-391.20	-570.50	-550.00	-506.70	-557.70	-522.58
99	-529.30	-557.10	-433.70	-494.90	-509.30	-560.90	-514.20
100	-594.80	-433.10	-657.80	-495.80	-614.70	-599.10	-565.88

Table C.18: The realistic MBIE agent with 5 reaches and 3 habitats per reach

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
1	-3,937.25	-2,826.65	-2,945.05	-2,848.85	-2,002.05	-2,911.97	687.66
2	-2,494.65	-3,033.15	-3,623.95	-4,788.35	-3,810.65	-3,550.15	864.26
3	-3,044.25	-4,003.05	-3,419.55	-1,681.85	-3,516.55	-3,133.05	880.33
4	-2,623.95	-3,070.35	-3,179.65	-3,646.45	-2,343.35	-2,972.75	505.97
5	-2,474.05	-3,661.75	-2,775.35	-3,647.15	-4,602.75	-3,432.21	839.43
6	-2,092.75	-2,297.95	-2,628.15	-3,057.05	-3,388.15	-2,692.81	533.00
7	-3,187.25	-2,718.75	-3,458.55	-3,082.05	-3,159.35	-3,121.19	266.05
8	-1,125.65	-3,074.75	-516.25	-4,052.15	-2,605.15	-2,274.79	1,442.31
9	-1,673.75	-3,759.65	-1,315.35	-3,394.95	-4,529.25	-2,934.59	1,382.68
10	-3,251.45	-816.55	-2,683.95	-2,216.65	-982.65	-1,990.25	1,062.54
11	-2,519.55	-978.25	-1,975.25	-1,819.95	-3,817.65	-2,222.13	1,049.29
12	-1,095.55	-2,097.35	-564.95	-657.15	-1,739.75	-1,230.95	670.94
13	-2,157.45	-2,128.05	-1,097.85	-3,317.45	-2,862.55	-2,312.67	843.49
14	-3,657.95	-1,220.45	-2,769.55	-809.15	-3,354.25	-2,362.27	1,279.12
15	-1,615.05	-1,123.55	-1,030.15	-834.25	-2,598.65	-1,440.33	708.55
16	-3,876.05	-1,609.05	-2,571.05	-2,876.55	-3,345.75	-2,855.69	853.94
17	-731.95	-1,772.75	-835.25	-1,247.05	-1,816.55	-1,280.71	507.43
18	-342.35	-1,072.35	-1,532.05	-713.65	-1,201.75	-972.43	458.18

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
19	-1,037.05	-1,950.35	-3,438.05	-1,841.95	-856.55	-1,824.79	1,021.85
20	-4,497.85	-1,985.85	-539.15	-627.85	-2,732.25	-2,076.59	1,640.37
21	-4,138.05	-2,848.95	-854.35	-2,933.25	-908.95	-2,336.71	1,422.93
22	-919.95	-3,749.65	-1,839.85	-1,523.05	-1,389.25	-1,884.35	1,093.97
23	-3,589.75	-2,159.85	-2,252.65	-2,507.55	-1,111.95	-2,324.35	885.57
24	-2,636.55	-689.55	-918.45	-4,147.75	-3,687.25	-2,415.91	1,572.18
25	-1,502.15	-1,210.55	-1,910.05	-669.65	-806.85	-1,219.85	507.31
26	-3,875.15	-3,734.05	-1,098.55	-1,195.15	-599.25	-2,100.43	1,572.83
27	-4,030.45	-821.05	-484.35	-1,190.95	-1,696.15	-1,644.59	1,407.58
28	-2,480.95	-688.85	-3,947.55	-3,329.75	-594.85	-2,208.39	1,522.25
29	-3,145.25	-588.95	-4,287.45	-2,421.65	-986.55	-2,285.97	1,527.34
30	-2,198.05	-1,147.15	-2,180.25	-1,511.75	-2,362.05	-1,879.85	523.53
31	-1,085.15	-929.95	-628.55	-853.05	-1,506.75	-1,000.69	327.31
32	-1,603.95	-1,432.15	-890.75	-1,861.25	-1,660.15	-1,489.65	368.17
33	-3,012.55	-617.55	-1,404.55	-2,118.95	-1,846.55	-1,800.03	884.41
34	-444.15	-4,046.85	-1,954.65	-1,507.95	-1,067.65	-1,804.25	1,372.24
35	-2,247.05	-792.15	-1,990.65	-1,841.75	-1,090.75	-1,592.47	620.77
36	-634.35	-870.45	-452.15	-1,921.45	-883.45	-952.37	570.51
37	-1,943.75	-1,221.95	-1,511.15	-2,959.75	-893.35	-1,705.99	800.17
38	-1,026.55	-3,338.75	-959.65	-708.15	-1,178.85	-1,442.39	1,073.63
39	-276.25	-2,110.05	-1,602.75	-854.45	-827.75	-1,134.25	721.12
40	-813.45	-721.25	-641.05	-552.05	-663.85	-678.33	97.01
41	-739.45	-1,564.35	-1,795.95	-1,073.05	-1,712.85	-1,377.13	453.40
42	-634.45	-421.85	-2,505.95	-543.65	-1,821.75	-1,185.53	928.33
43	-1,074.45	-871.25	-896.35	-1,852.25	-1,515.65	-1,241.99	427.87
44	-802.25	-1,219.05	-789.65	-809.45	-379.75	-800.03	296.82
45	-1,621.55	-747.55	-3,633.75	-1,491.65	-1,015.55	-1,702.01	1,136.28
46	-1,970.85	-784.55	-1,792.15	-1,275.15	-4,085.95	-1,981.73	1,264.54
47	-2,271.55	-1,221.35	-2,769.45	-1,669.55	-1,960.15	-1,978.41	587.48
48	-642.05	-3,184.25	-1,116.45	-773.15	-1,734.35	-1,490.05	1,037.14
49	-844.95	-670.25	-3,707.85	-1,319.75	-1,158.95	-1,540.35	1,238.23
50	-1,834.45	-471.55	-1,025.45	-627.65	-879.45	-967.71	530.14
51	-421.35	-2,373.25	-1,767.05	-587.25	-1,043.75	-1,238.53	821.10
52	-755.85	-2,100.65	-1,095.65	-516.55	-2,965.55	-1,486.85	1,023.74
53	-411.25	-2,006.65	-985.95	-1,475.95	-591.35	-1,094.23	653.70
54	-1,533.15	-660.85	-1,115.55	-1,807.45	-4,111.85	-1,845.77	1,338.99
55	-476.35	-840.55	-2,288.15	-1,529.75	-960.95	-1,219.15	707.28
56	-580.45	-439.35	-555.75	-2,531.65	-1,120.55	-1,045.55	871.47
57	-894.95	-641.75	-364.05	-1,024.65	-1,622.95	-909.67	472.15
58	-822.95	-717.45	-611.05	-836.75	-758.25	-749.29	91.21
59	-359.75	-763.85	-2,209.15	-709.15	-431.55	-894.69	754.98
60	-1,245.85	-581.65	-430.45	-1,296.75	-636.85	-838.31	402.82
61	-1,551.55	-1,542.45	-1,255.55	-581.25	-543.65	-1,094.89	500.59
62	-1,684.45	-617.75	-1,107.35	-1,343.35	-1,344.75	-1,219.53	394.30
63	-985.75	-2,057.45	-453.95	-786.45	-1,425.35	-1,141.79	621.00
64	-1,007.75	-536.45	-691.25	-3,065.15	-2,837.95	-1,627.71	1,223.02
65	-763.45	-1,401.85	-935.55	-1,345.85	-1,073.05	-1,103.95	270.42
66	-2,647.75	-1,066.85	-1,015.65	-1,387.85	-1,955.75	-1,614.77	688.11
67	-1,076.35	-581.15	-2,394.35	-1,209.05	-580.25	-1,168.23	742.27
68	-1,492.05	-912.25	-1,673.25	-755.85	-1,041.25	-1,174.93	390.96
69	-2,311.95	-649.85	-782.35	-2,112.65	-915.95	-1,354.55	791.79
70	-739.25	-1,066.55	-492.35	-852.75	-2,806.65	-1,191.51	926.34
71	-1,886.35	-1,992.05	-1,149.45	-1,041.45	-593.95	-1,332.65	592.75
72	-545.25	-808.85	-1,701.95	-859.25	-1,072.25	-997.51	436.20
73	-864.15	-901.95	-704.25	-762.75	-566.75	-759.97	133.66
74	-945.15	-579.15	-473.05	-2,263.65	-754.65	-1,003.13	727.11
75	-1,508.85	-2,947.45	-498.15	-687.95	-443.55	-1,217.19	1,057.65
76	-3,160.65	-2,004.65	-691.75	-2,252.35	-1,172.45	-1,856.37	962.29
77	-1,783.15	-767.55	-1,471.75	-1,346.15	-526.55	-1,179.03	518.08
78	-1,321.45	-1,397.45	-1,136.65	-1,266.85	-1,345.35	-1,293.55	99.48
79	-932.35	-807.85	-1,935.95	-1,242.45	-1,128.75	-1,209.47	439.74
80	-498.45	-1,537.85	-356.25	-1,099.55	-2,564.15	-1,211.25	892.83
81	-1,467.75	-1,730.05	-607.85	-1,358.15	-1,146.35	-1,262.03	421.88
82	-584.65	-1,688.55	-3,056.85	-1,203.55	-1,090.05	-1,524.73	942.08
83	-536.85	-969.95	-1,121.25	-893.95	-1,125.05	-929.41	240.87

	Reward per episode						
Episode	1	2	3	4	5	Mean	Std. dev
84	-1,926.35	-538.05	-468.35	-874.95	-1,402.55	-1,042.05	617.05
85	-1,009.95	-286.35	-2,825.65	-927.05	-1,835.15	-1,376.83	979.18
86	-3,171.15	-355.15	-1,018.75	-1,076.15	-724.45	-1,269.13	1,101.10
87	-544.35	-778.55	-2,990.95	-1,129.05	-2,248.75	-1,538.33	1,042.54
88	-989.75	-795.15	-782.25	-689.65	-1,178.25	-887.01	196.05
89	-4,120.95	-587.35	-1,312.65	-2,001.15	-1,104.35	-1,825.29	1,379.86
90	-938.35	-407.55	-498.85	-1,014.75	-1,711.25	-914.15	518.46
91	-1,237.75	-414.35	-971.55	-636.75	-726.95	-797.47	316.91
92	-614.15	-661.25	-519.45	-2,932.05	-1,602.05	-1,265.79	1,029.15
93	-815.75	-605.75	-1,186.45	-690.85	-521.65	-764.09	259.94
94	-562.85	-1,079.85	-1,630.95	-2,366.35	-3,449.55	-1,817.91	1,131.14
95	-639.75	-525.35	-513.85	-393.65	-1,437.15	-701.95	420.12
96	-1,165.05	-703.95	-2,595.45	-1,907.35	-1,118.65	-1,498.09	751.21
97	-501.15	-2,380.65	-958.15	-1,294.45	-2,773.95	-1,581.67	961.60
98	-374.15	-696.75	-902.85	-370.25	-1,198.65	-708.53	355.09
99	-3,221.95	-2,241.55	-447.05	-1,009.05	-3,550.65	-2,094.05	1,351.43
100	-2,539.55	-743.55	-478.95	-476.35	-790.65	-1,005.81	869.68