Abstract

This thesis examines a series of problems with the goal of better understanding the fundamental dilemma of whether to invest effort in obtaining information that may lead to better opportunities in the future versus exploiting immediately available opportunities. In particular this work investigates how this dilemma is affected by competition in an evolutionary setting. To achieve this requires both the use of evolutionary game theory, and Markov decision processes or stochastic dynamic programming. This thesis grows directly out of earlier work on the Social Learning Strategies Tournament. Although I cast the problem in the biological setting of optimal foraging theory, where it fills an obvious gap, this fundamental dilemma should also be of some interest to economists, operations researchers, as well as those working in ecology, evolution and behaviour.
Acknowledgments

Peter, who’s to say where this math adventure began: In a sunny courtyard at a conference in Victoria? In Watson Hall during a math and poetry lecture? In Jeffery 201 up at the board trying to explain inclusive fitness to you and Troy? Wherever and whenever it started, you were there. I’ve been amazed over the years, at your ability to always be there, right when I need you. I’ve checked with your other students, you have a gift. I feel honoured by your trust in letting me pursue my own thoughts and goals, however murky and vague they may have been at first. I am grateful for your willingness to sit down and talk me through my confusions, and your confidence that this will all lead somewhere at least interesting and complex, and with a little luck even beautiful. So thank you, for being a part of the happiest and hardest years of my life.

There are many people who have made my time as a student rich and full and joyous and they need to be thanked. Thank you Queen’s Math Bio Crew, past and present, the countless lab meetings, collaborations, dinner parties, afternoons of sledding, bike rides, classes, and assignments were all a delight because of the company. Special thanks to Amy for bringing me into the fold and Wes for seeing me out. Thank you thesis committee, for your careful reading, and your validation, in particular thank you Serdar for seeking me out and encouraging me to make my
thesis all it should be. Thank you to my favourite procrastination partners down in
the Jeffery Grad Dungeon, for indulging my various rants, and feeding me fuel for
more, not working with you was the best. Thank you Tim for showing me how to
work hard on a project and how to have a great time in the process, I’m looking
forward to a lifetime of collaboration and friendship with you.

There are many people who have made my time as a Kingstonian rich and full and
joyous, and they also need to be thanked, if only for being so understanding when I
abandoned them to work on this thesis. Thank you Kiel. Thank you Wendy. Thank
you Asia. There are the Applecrispers, and the Modern Fuelers, the Artelians and
the Fartelians. There is the smooth music appreciation choir of Kingston (SMACK),
and there is Izumo sushi and the people who ate there with me. I love you all and
the magic that you have created here in the ever changing mosaic of Kingston.

Thank you Mum and Dad, for always telling me I could do whatever I wanted. I
believed you. Thank you Andrea for listening to me blab on about math since you
were little. Thank you Claire for keeping me in the loop when I forget to be. Thank
you Ged and Ceal for your support.

There is one person without whom none of this could have ever happened. Bronwyn,
I know I have not been the easiest person to live with and love for this last year of
thesis. Thanks for making life sweet and good even when I was too far down a math
hole to notice. I am looking forward to making it up to you for the rest of our lives.
Statement of Originality

I certify that this thesis, and the research to which it refers, is my own work, and that any ideas or quotations from the works of other people, published or otherwise, are fully acknowledged in accordance with standard referencing practices.
## Table of Contents

Abstract i
Acknowledgments ii
Statement of Originality iv
Table of Contents v
List of Tables viii
List of Figures ix

1 Introduction 1
1.1 The Social Learning Strategies Tournament 1
1.2 The Rules of The Tournament 2
1.3 How to Learn 5
1.4 When to Learn 6
1.5 Polishing and Tweaking 11

Chapter 2: Background and Motivation 16
2.1 Optimal Foraging 16
2.2 Evolutionary Game Theory 20
2.3 Partially Observable Markov Decision Processes (POMDPs) 22
2.4 Motivating the problems of this thesis 26

Chapter 3: Duck Problem 1 31
3.1 Lone Forager, Free Switching,
Variable Terminal Time, Dynamic Environment 31
3.2 Many Foragers, Free Switching,
Variable Terminal Time, Dynamic Environment 46
B.4 Implementation of Replicator Dynamics and Dynamic Programming 214
B.5 Program for computing an approximation of $d_{source}$, $O_{source}$ and $O_{sink}$ 226
List of Tables

1  Optimal thresholds for when to switch, observe, and exploit  . . . . . .  81
2  The ES probabilities of employing each action  . . . . . . . . . . . . . . . . . .  102
# List of Figures

1. A contour plot of the approximately optimal number of rounds between checking in on the low quality patch, as a function of the high and low patch qualities. ................................................................. 32

2. The equilibrium proportion of foragers at the h patch as a function of q_o for the q_h values of 1, 5, 10, 20, 40, and 80. ................................................................. 47

3. c_o and c_s are not the costs of observing and switching but rather the factor by which payoff is reduced when these actions are employed. Thus the cost or loss due to observation from patch x is q_x(1 − c_o) and the loss or cost due to switching to patch y is q_y(1 − c_s), relative to already being at the y patch, and q_x − (q_y · c_s) relative to staying at the x patch. ................................................................. 78

4. The equilibrium type achieved and the round that it is achieved on as a function of the true value of q_y and the strategy of the population as manifest in d_x(t) and d_y(t). ................................................................. 99

5. Illustrations of how the foragers will respond to the quality of the source patch, h, changing to q'_h or the quality of the sink patch o changing to q'_o. q'_h is the critical value q'_h must exceed in order to instigate migration from the former sink patch to the former source patch. q'_h can be computed as the minimum q'_h satisfying d_h < 1 − d_{source}(q_o, q'_h). Similarly q'_o is the critical value which q'_o must fall below in order to instigate migration from the former sink patch to the former source patch. q'_o can be computed as the maximum q'_o satisfying d_h < 1 − d_{source}(q_o, q'_h) ................................................................. 111

6. The equilibrium proportion of foragers at the source patch as a function of the quality of the sink patch values for source patches of value 1, 5, 10, 20, 50, and 100. ................................................................. 136

7. Contour plot of the ES probability of observation at the source patch as a function of source and sink patch quality. ................................................................. 137

8. Contour plot of the ES probability of observation at the sink patch as a function of source and sink patch quality. ................................................................. 138
Chapter 1

Introduction

1.1 The Social Learning Strategies Tournament

In January 2008 members of our lab group received a forwarded email from a friend in the biology department. A few phrases in particular caught my eye.

“...a major international multi-disciplinary tournament on the evolution of social learning, inspired by Robert Axelrod’s famous Prisoner’s Dilemma tournaments on the evolution of cooperation...entrants will submit behavioural strategies detailing how to respond to the problem of resource gain in a complex, variable environment in terms of combinations of individual and social learning...the social learning strategies tournament will increase understanding of, and stimulate research on, the evolution of learning...”

My recently completed master’s project had been a first attempt at understanding how the optimal balance between investment in information against investment in exploiting that information is affected by frequency dependence of other agents facing the same problem. Questions of how and when to learn were foremost on my mind,
especially when this learning was happening in a social and hence game theoretic environment. For my PhD thesis I hoped to apply the tools of evolutionary game theory to learning strategies and this tournament seemed the ideal place to start. I also hoped for an opportunity to collaborate with a new friend of mine from the centre for neuroscience studies, Timothy Lillicrap. I suspected that his knowledge of machine learning and computer programming and my knowledge of game theory and mathematics would be a perfect combination for the task at hand. I asked my supervisor if I could take the next six months to focus on the tournament, and received the following warm encouragement “I have heard of this, and I thought of you at the time, but it came just as I was leaving. Yes yes yes, go for it. Tim would be an excellent collaborator.” Tim and I started work immediately.

The tournament’s goal was to answer the following question. What is the best way to learn in a complex, changing world? The answer to this questions was, like many worthwhile problems, “It’s tricky and it really depends on the specifics.”

### 1.2 The Rules of The Tournament

The specifics for this particular problem are as follows. The world of the tournament consists of 100 foragers, and 100 patches. Every round each forager must either specify a patch to forage at, or attempt to learn about a new patch. If the forager wants to learn about a new patch there are two ways it can go about this. The forager can observe, which entails selecting a model forager (or in some cases several model foragers) at random from the population of foragers currently exploiting patches. The observing forager becomes aware of the model forager’s patch and its payoff. Note that when observing, a forager may observe patches that it is already aware of. In
this case, although the forager does not learn about a new patch, it does potentially learn something new about the current payoff of a patch. The observation process is potentially error prone both with respect to actually learning about the right patch, and with respect to correctly assessing the payoff at that patch. If a forager innovates a patch is selected at random from those patches that the forager is not yet aware of and the forager becomes aware of this patch and its payoff, with no error. A forager must become aware of a patch either through observation or innovation before the forager can exploit that patch. If a forager spends the turn learning then it receives no payoff, if it spends the turn exploiting then it receives the payoff associated with that patch. Of critical importance is that the payoff of a patch depends solely on the quality of that patch, and not in any way on the number of foragers exploiting that patch.

Once the foragers make their moves and receive their payoffs, some of the foragers die. Each forager, independently, has a 1/50 chance of dying. From the remaining foragers, parents are chosen to reproduce and thus replace the dead foragers. A forager is chosen to be a parent with probability proportional to its average per round lifetime payoff, that is the total reward it has accrued over its lifetime divided by the number of rounds it has been alive. Usually the child is a clone of the parent, but occasionally the child is a mutant, using a strategy different from that of its parent.

After the births and deaths the patches are subject to a stochastic process which reassigns their qualities. Each patch changes, independently, with the same probability, \( P_c \). \( P_c \) is unknown to the foragers. When a patch changes its value is redrawn from some unknown distribution. This distribution is the same for all patches. The only information available about this payoff distribution is that it has relatively few
high payoffs and relatively many low payoffs, and all payoffs are positive.\footnote{Tim and I primarily used a geometric distribution in our simulations, with a wide range of parameters, as this is the simplest distribution meeting the given criteria. It turns out that this is also what the tournament organizers used, although we were careful to calibrate our entry to deal with more exotic distributions that also satisfied the given criteria.} Once this environmental change process is complete the round is over and the whole process begins again.

Tournament entrants were to submit formal rules, either in the form of pseudo code or matlab code, that specified which action a forager should choose each round, based on the history of the forager. The relevant information available to a forager was what their actions had been, what the identity and value of all the patches they had observed, innovated or exploited at were, and the payoffs that they had received, for every round of their life. A pair of strategies, \( A \) and \( B \) compete in the following way. A simulation consisting of 10000 rounds is run, with the initial population consisting entirely of one strategy type, say \( A \). After 100 rounds of the simulation mutation is enabled so that with a 1/50 chance the child of an \( A \) type is a \( B \) type and vice versa. For the last 2000 rounds of the simulation the frequencies of the strategy types on each round are counted and the average frequency over these 2000 rounds is a strategy’s score for that simulation. In the tournament, every strategy was paired up against every other strategy, round robin style, and competed in twenty simulations with each partner, ten as the initial resident population and ten as the invading mutant population. There were 104 entries in total, resulting in over 100,000 simulations to be conducted. The top ten scoring strategies from this first stage then went on to a second melee stage. Throughout the first stage of the tournament only a single set of environmental parameters was used. In the second stage of the tournament the top ten strategies from the first stage competed simultaneously against each other across
a wide variety of environmental parameters. The winner was the strategy with the highest mean score from this second melee stage. The tournament organizers reserved the right to run simulations until they were satisfied that there was a clear winner.

1.3 How to Learn

The first thing Tim and I did was implement the tournament’s formal process as a matlab program so that we could run our own mini-tournaments and test the performance of strategies. Once we had the formal process of the tournament up and running, we conducted weekly battles with each other. We each coded a strategy independently and then pitted these strategies against each other over a variety of environmental parameters. To keep motivation high the loser of our weekly battles made dessert for the winner. After each competition we would analyze the results, inspect the behvaiours of individual foragers to see if what they were doing was sensible, and discuss the strengths and weaknesses of each strategy and how the strategies might be improved. From these weekly competitions insights and intuitions emerged that would be key in the development of our final strategy.

The first of these insights was that learning through observation was in almost all cases better than learning through innovation. In retrospect this seems obvious, but it was not clear to us until after several months of thinking about the problem and playing with test strategies. If a forager wants to learn about a new patch, it can do so by either observing or innovating. We assume that a forager wants to learn about this new patch, not for the sake improving its estimates of the patch quality dynamics, but because it wants to learn about a patch that it can exploit, and it feels that the expected quality of the patches it currently knows about are too low. Thus the forager
would like the patch that it learns about to have the highest possible quality. Unless
the population that the forager exists in is full of pathologically bad strategies, the
patches that other foragers are exploiting at will on average have at least as good as if
not better quality than a patch chosen at random through innovation. Thus foragers
should almost always observe and almost never innovate. A detailed discussion of
this point can be found in appendix A.1

1.4 When to Learn

The question of how best to learn, in the case of this tournament could be broken
down into two separate questions of how and when to learn. The how, from our first
insight is clear, always observe, but the when of these observations does not have
such a clear cut solution. The critical choice made by a forager is between exploiting
the best patch currently available versus paying the cost of a round without payoff
in the hopes of observing a better patch. Whether to forgo immediate rewards for
potential future gains is a ubiquitous problem and depends critically on the way in
which future rewards are discounted.

To determine the appropriate discounting for the world of the tournament we
must consider the goal of a strategy: to maximize the frequency of its strategy type.
Since there is no social interaction between foragers in the tournament, a strategy
can achieve this goal best by simply having each individual maximize its individual
expected birth rate. If we let \( R_i \) be the reward or payoff to a forager on round \( i \) of
its life, and \( P_d \) be the probability of death, then under the death birth process of the
tournament the fitness of a forager is proportional to the following.
CHAPTER 1. INTRODUCTION

\[
F \sim \sum_{t=1}^{\infty} (1 - P_d)^t \sum_{i=1}^{t} \frac{R_i}{t} = \sum_{t=1}^{\infty} (1 - P_d)^t R_t \sum_{i=0}^{\infty} \frac{(1 - P_d)^i}{t + i} \tag{1}
\]

The relative weights given to early versus later rewards in this fitness function defines the correct discounting policy for rewards. Thus the discount factor for the reward received on round \(t\) is

\[
(1 - P_d)^t \sum_{i=0}^{\infty} \frac{(1 - P_d)^i}{t + i} \tag{2}
\]

Clearly then the appropriate discounting of future rewards in the case of the tournament is a blend of hyperbolic and geometric discounting. The geometric terms \((1 - P_d)^{t+i}\) give the probability that the forager will be alive on round \(t\) to receive the reward, and alive on subsequent round \(t + i\) to have this reward contribute to its probability of having a birth. The hyperbolic \(1/(t+i)\) factor is due to the tournament’s fecundity function which has a forager’s probability of birth as proportional to the lifetime average of the foragers per round payoff. Rewards received later in life have less of an impact on this average, and thus less of an impact on fecundity than rewards received early in life.

This would be the end of the discounting story if foragers were faced with a simple trade off between actual present rewards, and actual future rewards, but the choice faced by foragers in the tournament is not so straightforward. Foragers must choose between actual reward in the present and information that may lead to higher future rewards. The information gained from observation can help a forager both directly in that the forager may become aware of a higher quality patch, and indirectly in
that observation will likely improve some of the forager’s estimates of the simulation parameters. Improved estimates will in turn likely improve the forager’s future decisions, and this benefit lasts for as long as the forager lives. In contrast if the forager becomes aware of a high quality patch, this immediately and directly improves its reward intake rate, however this benefit only lasts so long as the high quality patch remains unchanged. Because the indirect value of observation is often small in comparison to the direct value of observation, and because the indirect value of observation can only be assessed by a computationally intensive Bayesian sequential decision theory approach, Tim and I chose to focus only on the direct value of observation information. While the discounting of the future value of rewards is totally accounted for in expression 2, the value of information needs to be further discounted by the probability that this information will become irrelevant and cease contributing to the forager’s fitness. Specifically this information should be discounted geometrically by the probability that a high quality patch discovered through observation does not change quality on a given round, \((1 - P_c)\), as this would render the information from observation irrelevant.

This idea of discounting the value of a decision based on that decision’s probability of becoming irrelevant is subtly different from some of the other reasons for discounting future rewards. Thinking about the effects of extreme values of \(P_c\) on the optimal behaviour of a forager helps to develop an intuition for the relevancy window of a decision to acquire information through observation. Suppose that \(P_c = 1\), this means that every round every patch has its quality redrawn from the payoff distribution, so each patch is essentially the same, and there is no point in moving from patch to patch looking for a good patch, because there is no correlation between the payoffs
at a given patch over time. In this case the information gained through observation becomes irrelevant as soon as it is acquired. If $P_c = 0$ then the environment is totally static and a patch’s value will never change. In this case the information gained from an observation, provided it is correct, never becomes outdated, thus there is strong incentive to observe. For intermediate values of $P_c$ there is an intermediate level of structure to the environment, in the form of correlation of the payoffs at a patch from round to round, and the information gained through observation remains relevant for some random amount of time dependent on the specific value of $P_c$.

With an understanding of the desired qualitative behaviour of an optimal rule for deciding when to observe and when to exploit we set about devising a formal rule for making such decisions. A key aspect of this task was taking the raw information of a forager’s lifetime and distilling this into a set of relevant statistics upon which the decision function would be based.

The ideal formal decision rule, is one derived from expanding an infinite horizon decision tree according to the principles of Bayesian sequential decision theory, and then using stochastic dynamic programming to determine the optimal decision policy. A decision tree cannot be expanded indefinitely, so a truncated version of the infinite decision tree is used. The further into the future this truncation point lies the closer the policy computed by this procedure will be to the optimal policy [4]. To expand this decision tree requires an estimation of the payoff distributions of all patches and observable patches, not only on the round on which the first decision is made, but also an estimation of what these estimates will be on future rounds contingent on all possible outcomes of the previous decision. As a result the number of possible states grows exponentially as decisions further into the future are considered. Thus
within a very short number of time steps it becomes computationally unfeasible to keep track of all possible states a forager might find itself in. Typically, as is the case here, the state space explosion, as this phenomenon is sometime referred to, will be such that the number of steps into the future that a decision tree can be expanded before over running the memory of even the most powerful computers, is no where near the number of steps into the future required to find a good approximation of the optimal policy.

In contrast to the ideal decision rule our decision rule used only an estimate of $P_c$, and estimates of the first moment of the payoff distribution, and the first moment of the observable patches’ payoff distribution, although even estimating these basic parameters proved to be a non-trivial inference problem, with many complications and subtleties.

In this simplified setting, our problem of when to exploit and when to observe was rephrased as: how much greater does the expected payoff of an observed patch need to be than the expected payoff at the best currently available patch in order for observation to be worthwhile. Ideally this question of how much better should take three factors into consideration. The time a forager will have at the newly observed patch to recoup its losses, the relative fitness value of future rewards compared to rewards in the present, and the improvements of the forager’s estimates of the relevant simulation parameters resulting from the new information gained from either action, which will in turn improve future decision-making ability, and thus fitness. The rule we used focused solely on how much time a forager had to recoup its losses. Thus future rewards were discounted geometrically by the probability of death and the probability of the patch that the forager was currently exploiting changing, but the
CHAPTER 1. INTRODUCTION

hyperbolic discounting due to the particulars of the fecundity function was ignored. We also ignored the collateral fitness value of improved parameter estimations.

Balanced against these drawbacks is the relative simplicity of our rule. The general motivation for the rule can be summarized as follows. A forager should observe if it thinks that the quality of the patch that it will see is large enough that it will earn more payoff exploiting at the observed patch, before it dies or that patch changes quality, then if it were to exploit at the best available patch until it dies or that patch changes quality, given that it will lose a turn of foraging when it observes.

1.5 Polishing and Tweaking

Using this simple decision rule our entry was able to handily out compete any of our earlier attempts, except in environments with a very slow rate of change (low $P_c$). In high $P_c$ worlds the fluctuations of the environment ensure a minimum accuracy for the estimates of the mean of the payoff distribution, and the mean payoff of observable patches, since the fluctuating environment guarantees that foragers will have some minimum number of data points to use in their estimations. In contrast in a low $P_c$ environment, foragers using our decision rule will often exploit at the first patch that they observe, until their death. This is a somewhat sensible policy, since the basic premise of our strategy is that other foragers are likely to be exploiting at high quality patches. The issue with this policy is that observation is error prone, and so while typically in low $P_c$ environments other foragers will be exploiting at high quality patches, because of observation error there is some risk in naively exploiting at the first patch ever encountered. This is a pitfall of our strategy failing to consider the indirect value of observation resulting from improved estimates of the environmental
parameters. In high $P_c$ environments this risk is small since patch values change frequently, but in low $P_c$ environments this risk is potentially devastating. To remedy this flaw we overrode our basic decision function so that if $P_c$ was estimated to be low, and the forager had been exploiting at the same patch uninterrupted for a long time, it would occasionally observe, just to double check that it was indeed exploiting at a patch of comparable quality to the rest of the population, and to ensure a minimum number of data points for estimation.

Our decision rule, thus far, evaluates the value of observation based on the estimated mean of payoffs associated with observation, however several simulation parameters, namely the number of foragers observed at once, and the degree of error in observation both in terms of how payoffs are distorted and the chances of learning about the wrong patch, were certain to affect the value of observation. As observation becomes more error prone the value of observing will surely decrease, but to what extent is not clear. As the number of foragers observed in a single observation increases, and hence the amount of information revealed by observation increases the value of observation will surely increase, but to what extent is again unclear. These factors interacted in non-linear ways to affect the value of observation but we did not know how incorporate them into the decision rule. To this end, Tim who is something of an expert on neural networks and machine learning proposed that we train a neural network to tweak the basic outcome of our core decision rule based on relevant statistics for these factors.

After two months of playing with neural networks, the submission deadline was fast approaching. We were at the point where our entry with the neural network was able trounce the network free version of itself in rapidly changing environments
when the number of foragers observed at once was high, but showed indistinguishable performance in mildly fluctuating environments, and seemed to be actually making worse decisions in slowly changing environments. The simple solution to this problem was to run more simulations in low and middle range $P_c$ environments and broaden the data set from which the neural network learned, but there was not enough time left to run these simulations. In the end we simply chose an arbitrary cutoff in terms of the estimated $P_c$ value for when our strategy would use the neural network, and when it would use its simple discounting rule.

After all of our extensive simulations and theorizing, our end product felt like something cobbled together with duct tape and wire. Despite our months of work, the number of parameters of our strategy chosen by intuition, instead of statistically significant data and theoretically motivated principles, was more than either Tim or I were comfortable with. But the deadline was upon us, and the performance of our entry against everything else we had ever tried was very good, so on June 30th, 2008 we submitted Discountmachine and hoped for the best. There were 104 entries in all from sixteen different countries submitted by teams of academics from a wide variety of disciplines. We called our entry Discountmachine since the core decision of when to observe versus when to exploit was based on discounting future rewards, and then tweaking the result using a machine learned function.

In November of 2008 it was announced that we were in the top ten of the first round and so would proceed to the second and final round of the tournament. In March 2009 Tim and I were contacted by the tournament organizers and told we had won the tournament, beating out all other competitors by an impressive margin. We were invited to a conference in May of 2009 where the results of the tournament
were announced. At the conference we began collaborating with the tournament organizers on a paper describing the results of the tournament. In April of 2010 this paper “Why Copy Others? Insights from the Social Learning Strategies Tournament” [20] appeared in Science as a research article. As the title of the paper suggests the tournament organizers chose to focus on the simple and accessible insight of our strategy, that in the world of the tournament, innovation was useless and copying was the only game in town. Although Tim and I had found the problem of how to learn simple in comparison with the problem of when to learn, I am sympathetic to the organizers reasons for focusing on the simple insight. Though obvious in retrospect the result that innovation was useless came as a surprise to the tournament organizers and to all but a few of the 103 other tournament entrants. Of the top ten strategies the second place entrant was the only other strategy to exclusively observe. This result had a clear explanation, and a certain appeal in terms of accessibility. In contrast the decision rule that we had developed for when to learn was ad hoc, complicated, confusing, and not provably optimal, so it was difficult to draw any conclusions from its winning performance, other than that what it was doing was closer to the right thing than anybody else.

Winning the tournament, having the results published in Science, and the ensuing attention from popular media was very satisfying, but the important questions of how and when to learn raised by the tournament were left largely unanswered by our efforts and the tournament’s results. Particularly unsatisfying was that although this was an evolutionary tournament, in the development of our winning strategy we were able to completely ignore any game theoretic considerations. To that end I have spent much of the remainder of my time as a PhD candidate here at Queen’s working on
generalizations and simplifications of the formal problem posed by the tournament. In chapter two I outline the body of previous work that this thesis fits within, and introduce and motivate the general problem considered by the thesis. In chapters three, four and five I present a series of instances of the general problem outlined in chapter two. These problems can be seen as both a simplification and a refinement of the tournament’s formal problem. In chapter six I discuss how this work sheds some insight on the problem of how and when to learn, and the way this problem changes in a social and competitive setting.
Chapter 2

Background and Motivation

2.1 Optimal Foraging

I think of the tournament problem, and the problems investigated in this thesis as foraging problems and so I will primarily situate this work within the optimal foraging literature. Before going into optimal foraging theory, however it should be noted that this thesis, and the tournament problem, appear to lie at the heart of a broad and current trend in ecology, evolution and behaviour, characterized by explicitly considering the flow of information between individual organisms, and the evolutionary and ecological consequences of this information flow. “The ecology of information: an overview on the ecological significance of making informed decisions” [22] and “Information and its use by animals in evolutionary ecology” [7] serve as both reviews of and rallying calls for this approach. Information is of course relevant not only for foraging behaviours, but all realms of behaviour, specifically mate and habitat search, predator avoidance and vigilance, and reproductive timing. All of these behaviours can potentially be both frequency dependent in their effect on an
individual and involve trading off between immediate rewards and information that may or may not lead to greater rewards in the future. While the problems of this thesis could potentially been framed in any number of ways, I find the context of optimal foraging to be the most natural one.

Optimal foraging theory is based on the notion that evolution optimizes fitness, and that with regards to foraging behaviours, average energy intake rate is a good proxy for fitness. Within this broad framework of optimal foraging I find that there are two branches of particular interest. One branch of the theory focuses on finding the best learning and foraging strategy a lone forager can adopt. Papers in this lone forager branch typically specify a class of exploration exploitation trade off problems built up from specific biological assumptions and then use an optimization technique, usually Bayesian decision theory, to identify the best learning and foraging strategy for this class of problems. This best strategy can then be used as a benchmark to compare experimentally observed behaviours against. In the best case this provides insights into the cognitive constraints on learning behaviours and also insights into how experimental setups may differ from the natural environment to which the animals in question have presumably adapted their learning strategies. The other branch of interest focuses on understanding the distribution of a population of many foragers within an environment. Papers in the many foragers branch typically start by specifying the learning and foraging rules of an individual as given, and then use game theory to find the equilibrium distribution of foragers throughout the environment. Shockingly I have found only one paper in the optimal foraging literature, Olsson and Brown’s “Smart, smarter, smartest: foraging information states and coexistence” [19], which attempts to simultaneously consider both the optimal learning
and foraging rule and the population distribution that this learning and foraging rule creates. This paper is discussed in some detail later.

Problems from both branches take as given a patchy environment, typically discrete though there are exceptions [1], with patches of varying quality. The problems in either branch can be further separated into two broad classes of many patch and few patch problems. In many patch problems, foragers move through an environment consisting of a large, effectively infinite, number of patches so that patches are revisited with vanishing probability. In these many patch problems there is some distribution of quality over the patches, a forager depletes the quality of the patch it is foraging at, and the solution of these problems is when and under what conditions an optimal forager should stop foraging at one patch and move on to the next. See [18] for a typical single forager example, and [5] for a typical many forager example. Sometimes it is assumed that a forager knows the global distribution of patch qualities and must learn about the quality of its current patch to inform its decision to stay or go [18], although sometimes it is assumed that the forager can instantaneously assess the quality of its current patch, and it is the global patch quality distribution that is unknown and must learned [17]. Recently [8] and [9] have tackled the difficult case of learning simultaneously about the current patch quality and the global patch quality distribution. In few patch problems a forager must choose between some small number of patches, patch quality is not depleted by foraging, and the solution of these problems is when and under what conditions a forager should stop investigating the various patch qualities, i.e. exploring, and commit to a single patch, i.e. exploiting. In this situation, an optimal forager must weigh the cost of exploring other patches against the value and likelihood of detecting a patch of higher quality than its current
patch. See [14] or [23] for typical lone forager problems. Problems of this sort are referred to as multi-armed bandit problems, (after slot machines, the one armed bandit) or simply bandit problems. Bandit problems, first introduced to study the sequential design of experiments [3] and [21], have become the canonical model for studying exploration exploitation trade offs. When temporal heterogeneity is incorporated into a bandit problem so that the quality of each patch changes stochastically in time, the problem is then referred to as a wandering bandit problem. While stationary bandit problems are essentially optimal stopping problems, and have a general solution [10], in wandering bandit problems the accuracy of information decays and so a forager must always reevaluate its situation in light of possible changes in quality at other patches. Wandering bandit problems have received extensive study, not just in optimal foraging [25] [24], but also in machine learning,[26] [13] [2], and more recently in neuroscience [6]. Unlike the stationary bandit problem no general solution is known for the wandering bandit problem. Just as in the many patch problems there are aspects of the environment that could be taken as known by the foragers priori or that could be learned by the foragers. Even in the simplest case where it is assumed that foragers can instantly assess patch quality and that the foragers know the underlying dynamics of the environment the problem of how frequently to sample is still non-trivial. Knowing the stochastic dynamics of the patch quality is not the same as knowing the current patch qualities, thus if information is costly the essential exploration exploitation dilemma still exists even in this simplest case.

In particular the work of my thesis brings together the many and lone forager branches of optimal foraging theory in the few patches case. This seems particularly timely since quite recently Olsson and Brown’s “Smart, Smarter, Smartest: foraging
information states and coexistence” [19], has brought together the lone and many forager branches of optimal foraging theory in the many patches case. Their paper considers a population of many foragers living in an environment with a large number of patches per forager. Patch quality is depleted by foraging, but there is a slow, stochastic, and steady renewal process which balances this depletion. The paper considers three different classes of learning and foraging strategies, each requiring a different degree of cognitive or sensory sophistication, and finds for each class the Evolutionarily Stable (ES) strategy. (Olsson and Brown use the term optimal strategy which is not technically correct. The strategies they find are best described as ES strategies, since these strategies are the best responses to the equilibrium patch quality distributions generated when the ES strategies are employed by the population as a whole.) The paper goes on to compare the relative advantages of each class of learning and foraging strategy. They then explore the interesting possibility of coexistence of different species exploiting the same patchy resources, as a result of asymmetries in the costs and value of information resulting from the different patch quality distributions generated by the different strategy types.

### 2.2 Evolutionary Game Theory

Recall from chapter 1 that when addressing the core dilemma of the tournament, the interactions between foragers could be treated as part of the background environmental dynamics. Thus while the foragers were competing to reproduce, the form of this competition was more like that of two people playing solitaire side by side, in contrast to playing poker with each other. If the social interaction between foragers had been more substantial, then in addition to Bayesian or Markov decision theory,
the tools of evolutionary game theory would also have been required to tackle the
tournament problem. Intrigued by this possibility I began to search for simple prob-
lems whose solutions would require both evolutionary game theory and Bayesian or
Markov decision theory.

The core idea of evolutionary game theory, the evolutionarily stable strategy (ESS)
[16], relates to the general many forager optimal foraging problems in the following
way. The general premise of many forager problems is that given a patchy environ-
ment being exploited by a population of foragers, the learning and foraging strategy
adopted by the population as a whole will create a specific patch quality distribu-
tion in space and time. For that specific patch quality distribution there will be a
learning and foraging rule which is optimal for that patch quality distribution. Most
strategies will not in general be the optimal response to the patch distributions that
they create when employed by the population as a whole. In fact strategies that are
the best response to the environment that they produce are typically rare and given
the special term of Evolutionarily Stable Strategy. Given that the frequency of the
strategies employed by the individuals of the population are governed by a replicator
dynamic [27], the key idea is that in any population employing a strategy that is not
an ESS, it is possible for a rare mutant employing a strategy different from the resi-
dent population’s strategy, to replicate itself at a higher rate than the foragers using
the resident strategy. In this case it is said that the mutant can invade the resident
population. The Evolutionary stability of a population and its strategy profile is then
characterized as being uninvasive by rare mutants. Thus the notion of an optimal
learning and foraging strategy is only sensible in the lone forager case. When there
is social interaction between conspecifics notions of optimality must be replaced with
evolutionary stability.

2.3 Partially Observable Markov Decision Processes (POMDPs)

Some of the problems in this thesis are best described formally using the language of Partially Observable Markov Decision Processes. Before proceeding with the problems I give a quick overview of the basic notation and concepts of this theory.

Before formally defining a POMDP we first describe the basic Markov Decision Processes (MDP).

An MDP is a four tuple consisting of:

- A state space $S$.
- A set of possible actions $A$. Sometimes the possible actions may be restricted by the state in which case we have for each state $s \in S$ a set of possible actions $A_s$, with $A = \bigcup_{s \in S} A_s$.
- A state transition kernel $T_a(s, s')$ giving the probability of the next state being $s' \in S$ given that the current state is $s \in S$ and that the decision maker employs action $a \in A_s$ in the current time step.
- A real valued reward function $R_a(s, s')$, which assigns a reward to the decision maker for taking action $a$ when in state $s$, and having the system then transition into state $s'$.

The problem is to find a policy, in the form of a function $\pi : S \to A$, which prescribes which (legal) action a decision maker should take for all possible states.
The goal of a policy is typically to maximize the expectation of some function of the rewards over a period of time. This expectation of a function of the rewards, is referred to as the objective function. A standard objective function, that we will be using, is one of the form.

\[
E \left[ \sum_{t=1}^{\infty} (1 - P_d)^t R_{a_t}(s_t, s_{t+1}) | a_t = \pi(s) \right]
\]

(3)

or equivalently

\[
E \left[ \sum_{t=1}^{T} R_{a_t}(s_t, s_{t+1}) | a_t = \pi(s) \right]
\]

(4)

where \( T \) is geometrically distributed with parameter \( P_d \) (\( P \) for probability \( d \) for done).

Note that in this framework it is assumed that the decision maker knows everything about \( S, A, T, \) and \( R \), perfectly. The much more difficult problem, where there is uncertainty concerning \( S, A, T, \) and \( R \), can be approached using the techniques of reinforcement learning [26], which though interesting I will not go into here.

If the state space is finite then a technique known as dynamic programming can be used to compute an optimal policy \( \pi^* \) which maximizes the given cost function. Dynamic programming works in the following way. The key concept is the notion of a value function \( V_\pi(s) \) which gives the total expected future payoff given that policy \( \pi \) is employed. If the state space is finite and there is some discount factor in effect, in our case this discount factor is \( 1 - P_d \), then the values of each state can be computed by solving the following system of equations, where each equation is of the form.

\[
V_\pi(s) = (1 - P_d) \left( \sum_{s' \in S} (R_{\pi(s)}(s, s') + V_\pi(s'))(T_{\pi(s)}(s, s')) \right)
\]

(5)

This works precisely because once a policy has been specified, the Markov Decision Process behaves like a Markov chain. Once this system of equations has been solved and the value of each state is known under some given policy \( \pi_0 \) this policy can be
strictly improved by iterating through the states and for each state changing the policy so that the action used is the action which results in the highest expected value, under the current valuation of the states. Essentially $\pi_0(s)$ is changed to $\pi_1(s)$ where $\pi_1(s)$ is chosen to maximize

$$\sum_{s' \in S} (R_{\pi_1(s)}(s, s') + V_{\pi_0}(s'))(T_{\pi_1(s)}(s, s'))$$

(6)

This process is repeated, so that the values of each state under policy $\pi_1$ are computed and these values are then used to compute an improved policy $\pi_2$, and so on, eventually the policy produced by this process will converge on the optimal policy $\pi^*$, and the valuations of the states will converge to their true value under the optimal policy, typically denoted as $V(s)$.

The Partially Observable Markov Decision Process extends the basic MDP. In addition to the state space, $S$, the action space $A = \bigcup_{s \in S} A_s$, the transition kernel $T_a(s, s')$, and the reward function, $R_a(s, s')$, we also have the following. There is a set of possible observations $O$, and an observation function $\Omega_a(s, s', o)$ which gives the probability of the decision maker experiencing observation $o \in O$ given that action $a$ was used when the environment was in state $s$ and then transitioned to state $s'$.

Typically $o$ does not provide enough information for the decision maker to know with certainty the true state of the environment. In fact in order to make the best guess possible about the true state of the environment it will often be necessary for the decision maker to consider its entire history of observations. Thus the decision maker maintains a probability distribution over the possible states of the system. This distributions is referred to as the belief state of the decision maker and denoted as $b$, a possible belief state in the set of all possible belief states $B$, with $b(s)$ giving the probability that the current true state of the system is $s$. 
It is assumed that the decision maker knows $S$, $A$, $T$, $O$, and $\Omega$. Given that the decision maker’s current belief state is $b$ and that the decision maker takes action $a$ and observes $o$, the belief state of the decision maker is updated according to Bayes theorem, that is

$$b'(s') = z \sum_{s \in S} \Omega_a(s, s', o) T_a(s, s') b(s)$$

(7)

Where $z$ is a normalizing constant with

$$z = \frac{1}{\sum_{s' \in S} \sum_{s \in S} \Omega_a(s, s', o) T_a(s, s') b(s)}$$

This belief state updating process implicitly defines a belief state transition kernel, $\tau_a(b, b')$ which denotes the probability of transitioning from belief state $b$ to $b'$, given that action $a$ is taken. This can be computed as the probability of making an observation $o$ which would result in being in belief state $b'$ given that the decision maker is in belief state $b$ and takes action $a$.

Thus a POMDP can be reframed as a belief MDP, where the belief states of the POMDP become the states of the belief MDP, and the belief transition kernel $\tau$ becomes the transition kernel of the belief MDP. The reward function $R_a(s, s')$ of the POMDP, is used to define the expected reward

$$r_a(b, b') = \sum_{s \in S} \sum_{s' \in S} b(s) b'(s') R_a(s, s')$$

(8)

of being in belief state $b$ and taking action $a$. The action space is the same for the POMDP as it is for the belief MDP.

The problem is to find a policy, in the form of a function $\pi : B \rightarrow A$ which prescribes which action should be taken in each possible belief state. The goal of the
policy is to optimize some function of the expected reward. Typically

\[ E \left[ \sum_{t=1}^{\infty} (1 - P_d)^t r_{a_t}(b_t, b_{t+1}) | a_t = \pi(b_t) \right] \]  

(9)

or equivalently

\[ E \left[ \sum_{t=1}^{T} r_{a_t}(b_t, b_{t+1}) | a_t = \pi(b_t) \right] \]  

(10)

with \(T\) geometrically distributed with parameter \(P_d\).

When this approach is taken even if the state space of the POMDP is finite, the belief state space of the belief MDP will typically be infinite, and so the belief MDP cannot be solved by the simple policy iteration process described earlier.

### 2.4 Motivating the problems of this thesis

This thesis is primarily a response to the formal problem presented in the Social Learning Strategies Tournament. The tournament problem was too simple in that it lacked frequency dependence in the patch payoffs, and as a result the problem of whether to learn through observation or through innovation had, an approximately optimal, simple solution of always observe. However, the problem was too difficult in that it involved the simultaneous learning of both the current state of a dynamic and stochastic environment, and learning of the underlying stochastic dynamic processes of the environment. Thus it combined a difficult statistical problem of inferring from various series of data points the underlying process generating those series, with a difficult optimization problem of choosing which types of data points to sample given that the costs and benefits of different types of sampling are themselves contingent on the very stochastic processes to be estimated.

In the chapters that follow, I present a sequence of problems and their solutions
(or approximate solutions) which both incorporate frequency dependence into the
tournament problem while simplifying the problem in other ways to make it tractable.
Recall from chapter 1 that the critical dilemma of the tournament reduced to a
question of when to observe in hopes of finding something better versus going with the
best available option on hand and exploiting. Throughout this series of modifications
I seek to retain the essence of this particular exploration exploitation dilemma. The
goal being to develop an understanding of how frequency dependence and competition
affect the value of information and hence the solution of exploration exploitation
dilemmas of the type presented in the tournament.

The inspiration for how to modify the tournament problem comes from an early
ideal free distribution experiment [12]. In this experiment, two experimenters stand-
ing 20 meters apart distributed uniformly sized chunks of bread at various frequencies
to a flock of 33 ducks. They then observed that very quickly and without sampling
the ducks distributed themselves in such a way that the bread per duck at each patch
was very close to equal. This distribution of ducks, sometimes called the ideal free
distribution, is predicted by assuming that foragers are omniscient, migration has
negligible cost, and that foragers compete equitably for resources. If the assumption
of omniscience is removed and the ducks must pay some cost to learn about the bread
rate at the other patch then the problem of how much effort to invest in keeping track
of the quality of the other patch, given that the patch qualities are subject to change
is a relatively simple problem that captures the exploration exploitation dilemma of
the tournament, while incorporating frequency dependence.

The tournament environment consists of 100 patches, following the ideal free dis-
tribution experiment, the problems I will look at have a two-patch environment.
Foragers in the tournament environment were not initially aware of the existence of patches and had to spend time discovering a patch before they could exploit it. In contrast, given that there are only two patches in the following problems, foragers are assumed to be innately aware of the existence, though not necessarily the quality, of both patches.

As in the ideal free distribution model, in problems where there is a population of foragers payoffs are frequency dependent. Foragers are assumed to be equally competitive so that the payoff per round to a forager is the quality of the patch that the forager is currently exploiting at, divided by the proportion of the population also exploiting at that same patch.

In the tournament the fitness function was such that early rewards were valued more highly than later rewards introducing an analytically intractable hyperbolic discount factor. In the problems that follow the game is played out over the course of a single foraging session, as in the ideal free distribution experiment, and not over a forager’s life time. In this situation it is reasonable to assume that fitness is simply proportional to the total amount of reward accrued over the foraging session, avoiding the need for hyperbolic discounting. This shift in timescales means that the role of the probability of death, $P_d$, in the tournament is now replaced by the probability that the foraging session will end, which I continue to denote as $P_d$.

The environmental change process in the following problems, when present, is similar to that used in the tournament, except that for analytic ease, only one patch is allowed to change at a time. Thus each round with probability $P_c$ one of the patches has its quality redrawn from the payoff distribution, with each patch being equally likely to be the one that changes. For analytic tractability and in keeping with the
tournament, I assume that the payoff distribution is geometric, with positive support, and parameter $\lambda$.

In the tournament, both the rate at which patch qualities changed, $P_c$, and the quality distribution from which new patch qualities were drawn, were unknown to the foragers. I contrast I assume that the dynamics of the environment are innately known to the foragers. While this assumption is strong there are several reasons for employing it. I am searching for a simplest analytically tractable problem which retains the exploration exploitation dilemma at the heart of the tournament. Having foragers base their decisions on the true values of these parameters rather than estimates greatly simplifies the problem, while preserving this critical dilemma. The second reason is that in the following problems the foragers can do very little to affect the rate at which new data points for estimating $P_c$ and the quality distribution become available. Having better information about the true value of $P_c$ and the true quality distribution will likely enable foragers to make better decisions, but there is little that the foragers can do to improve these estimates other than wait. The third reason is that there are situations where this assumption is plausible. If the environmental dynamics have been stable over a long period of evolutionary time then it is possible that the genome of the forager population has encoded these dynamics and if the environmental dynamics have been stable over the course of a forager’s life time, and the situation described above is a single foraging session in a life time of such foraging sessions, it is plausible that the foragers’ learned estimates of $P_c$ and the payoff distribution have stabilized near their true values. In this latter case the optimal or evolutionarily stable foraging rule that we find based on the assumption of innate knowledge of $P_c$ and the quality distribution will be a good approximation
for the true optimal strategy in as much as the young foragers still learning these environmental parameters form an insignificant fraction of the total population.

In the coming chapters I will present three specific instances of the general and informally defined problem outlined above. In two of these three instances I will present both a lone forager optimization problem and an analogous evolutionary population game problem, and in one of these instance I will only present an evolutionary game problem. By comparing the value of information in a lone forager’s optimal strategy, to the value of information in the evolutionarily stable strategy of a population of foragers I hope to better understand how frequency dependence and competition affect the value of information.
Chapter 3

Duck Problem 1

3.1 Lone Forager, Free Switching,
Variable Terminal Time, Dynamic Environment

This section addresses the first specific instance of the general problem discussed at the end of chapter 2. We will first consider a single forager in this dynamic two patch environment, and in the following section we will consider the case where there is a large population of foragers in this environment. A forager can move freely between the patches, (as was the case in the tournament). A forager would like to always exploit at the patch that has the highest quality, however this is not possible since the forager knows only the current quality of the patch it has just foraged at, the quality of the other patch the last time it foraged there, and the time since it last foraged at this other patch.

Suppose a forager is at a patch of quality 15, and that the last time this forager knew the value of the other patch it had quality 7. Clearly this forager will spend
Figure 1: A contour plot of the approximately optimal number of rounds between checking in on the low quality patch, as a function of the high and low patch qualities.

most of its time foraging at the high quality patch. The dilemma faced by the forager is one of how frequently, if ever, to exploit at the low quality patch and check if it has changed. Two factors influence how much checking should be done: the cost of checking, which is the difference between the two patch qualities in this case 8 (assuming that there has not been a patch quality change), and the risk of the patch with quality 7 changing and having a new quality greater than 15, which depends on the probability of a quality change occurring, the distribution new patch qualities are drawn from, and the current value of the high quality patch. If the rate of environmental change is low and if there is a low probability of a new patch quality
being greater than the current high quality patch, there should be less checking. If the cost of checking is low then there should be more checking. An optimal foraging strategy is one that perfectly balances these risks and costs, maximizing the expected reward a forager will receive over the course of a foraging session. In this case of patches with quality 15 and 7, figure 1 shows that according to the approximately optimal policy calculated in the following sections, a forager should primarily exploit at the patch of quality 15, but that after every 18 rounds of exploiting at this high quality patch the forager should switch to the patch that formerly was and maybe still is of quality 7 to see if it has improved.

**Problem 3.1 Summary**

- Each round there is a patch quality change with probability $P_c$. Each patch is equally likely to be the patch that changes.
- If a patch quality changes, its new value is drawn from a geometric distribution with parameter $\lambda$.
- Each round, with probability $P_d$, the foraging session ends.
- There is no explicit cost to switching from one patch to the other, only implicit opportunity costs.
- A forager knows
  - The quality of the patch it is currently at.
  - The quality of the patch it is not currently at, the last time it was at that patch.
  - How much time has passed since it last knew the quality of the other patch with certainty.
  - The parameters of the model, $P_c, P_d$ and $\lambda$.
- A forager does not know
  - The current quality of the patch it is not at.

**What we are looking for:** The frequency with which a forager should check in on the lower quality of the two available patches, as a function of both the high and low patch qualities.
The problem can be formally described in the language of POMDP’s as follows.

### 3.1.1 POMDP formulation

The state space $S$ consists of the current quality of the two patches, thus $s \in S$ is of the form $s = (q_h, q_o)$, where $q_h, q_o \in \mathbb{Z}^+$ are qualities of the patch that the forager has most recently foraged at, $h$ for here, and the quality of the patch that the forager has foraged at some time ago, $o$ for other. The action space, $A$, consists only of whether to forage at the same patch as the round before, $a = h$ or to forage at the other patch, $a = o$.

The state transitions, $T_a(s, s')$, can be defined as follows.

\[
T_h((q_h, q_o), (q'_h, q'_o)) = 1 - P_c + (P_c/2)\lambda(1 - \lambda)^{q_h-1} + P_c/2\lambda(1 - \lambda)^{q_o-1}
\]

if $q_h = q'_h$ and $q_o = q'_o$ \hspace{1cm} (11)

\[
T_h((q_h, q_o), (q'_h, q'_o)) = (P_c/2)\lambda(1 - \lambda)^{q'_h-1}
\]

if $q_h \neq q'_h$ and $q_o = q'_o$ \hspace{1cm} (12)

\[
T_h((q_h, q_o), (q'_h, q'_o)) = (P_c/2)\lambda(1 - \lambda)^{q'_o-1}
\]

if $q_h = q'_h$ and $q_o \neq q'_o$ \hspace{1cm} (13)

\[
T_o((q_h, q_o), (q'_o, q'_h)) = 1 - P_c + (P_c/2)\lambda(1 - \lambda)^{q'_h-1} + P_c/2\lambda(1 - \lambda)^{q_o-1}
\]

if $q_h = q'_h$ and $q_o = q'_o$ \hspace{1cm} (14)

\[
T_o((q_h, q_o), (q'_o, q'_h)) = (P_c/2)\lambda(1 - \lambda)^{q'_o-1}
\]

if $q_h = q'_h$ and $q_o \neq q'_o$ \hspace{1cm} (15)

\[
T_o((q_h, q_o), (q'_o, q'_h)) = (P_c/2)\lambda(1 - \lambda)^{q'_h-1}
\]

if $q_h \neq q'_h$ and $q_o = q'_o$ \hspace{1cm} (16)
The reward function is simply the quality of the patch that the forager chooses to forage at, thus

\[ R_a((q_h, q_o), (q'_h, q'_o)) = q'_h \]

A policy \( \pi \) is a function which maps the information available to the forager to an action. At each time step the forager observes the value of the patch it has just foraged at and whether or not this is the patch that it foraged at on the previous round. The observation set \( \mathcal{O} \) is then \( \mathcal{A} \times \mathbb{Z}^+ \). Since there is no noise on this observation, the information a forager receives when they take action \( a \) and transition from state \( s = (q_h, q_o) \) to state \( s' = (q'_h, q'_o) \) is \( o = (a, q'_h) \) with probability one. The information available to the forager is its entire history of observations.

We assume that the forager has full knowledge of the dynamics of the patch qualities, thus the beliefs of the forager about the true state of the system, \( s = (q_h, q_o) \), depend only how many rounds have passed since the forager last knew the quality of each patch with certainty, and how many of these rounds it was possible for the \( q_o \) quality patch to have changed. Thus the forager’s belief distribution, \( b \in \mathcal{B} \), over the possible system states is most easily represented by the tuple \( b = (q_h, \hat{q}_o, t) \). Where \( q_h \) is the quality of the patch the forager has just foraged at, \( \hat{q}_o \) is the quality of the other patch when the forager was last there, and \( t \) is the number of rounds where there has been an opportunity for a quality change at the other patch since the forager last knew the quality of the other patch with certainty. Note that on rounds where the forager’s current patch has changed quality, the other patch can not have changed. From this we can compute the forager’s belief that the system is in a given state \( s = (q_h, q_o) \). We let \( b(s) \) denote the probability that the true state of the system is \( s \) given the current belief state \( b \).
Pr(s = (qh, qo)|b = (qh, ˆqo, t)) = (1 − P′ c)t + (1 − (1 − P′ c)t)(λ(1 − λ)qo)

if qh = qh and qo = ˆqo

Pr(s = (qh, qo)|b = (qh, ˆqo, t)) = (1 − (1 − P′ c)t)(λ(1 − λ)qo−1)

if qh = qh and qo ̸= ˆqo

Pr(s = (qh, qo)|b = (qh, ˆqo, t)) = 0

if qh ̸= qh

Where

\[ P′_c = \frac{P_c}{2 - P_c(1 - \lambda(1 - \lambda)qo-1)} \]

denotes the per round probability of the qo quality patch changing, given that a forager at the qh quality patch did not notice a change in the qh quality patch that round. Note that this gives the foragers beliefs about the current state of the system, but the forager is really interested in the state of the system on the next round since that determines the value of the actions that it can take.

The belief transition probabilities, τa(b, b′), gives the probability that the next belief state is b′ given that the current belief state is b and that the forager takes
action $a = \pi(b)$. $\tau$ can be defined as follows.

$$
\tau_h((q_h, \hat{q}_o, t), (q_h, \hat{q}_o, t + 1)) = (1 - P_c/2)
+ (1 - (1 - P_c/2))(\lambda(1 - \lambda)^{q_h - 1})
$$

(21)

$$
\tau_h((q_h, \hat{q}_o, t), (q_h', \hat{q}_o, t)) = (1 - (1 - P_c/2))(\lambda(1 - \lambda)^{q_h' - 1})
$$

(22)

$$
\tau_o((q_h, \hat{q}_o, t), (\hat{q}_o, q_h, 1)) = (1 - P_c/2)(1 - P_c')^t
+ (1 - (1 - P_c/2)(1 - P_c')^t)(\lambda(1 - \lambda)^{\hat{q}_o - 1})
$$

(23)

$$
\tau_o((q_h, \hat{q}_o, t), (\hat{q}_o', q_h, 1)) = (1 - (1 - P_c/2)(1 - P_c')^t)(\lambda(1 - \lambda)^{\hat{q}_o' - 1})
$$

(24)

When a forager switches to the other patch and finds it changed, the forager does not know whether or not this patch has changed just this round or earlier. If this other patch has changed just this round it is not possible for the forager’s former patch to have changed on that same round, and so in this case it would be appropriate for the $(\hat{q}_o', q_h, 1)$ term in equation 24 to be $(\hat{q}_o', q_h, 0)$. The probability of this other patch having just changed versus having changed some time in the past is dependent on when this other patch was last visited by the forager. For analytic simplicity we assume, that the forager always assumes, that the other patch changed some time in the past, regardless of how likely this actually is. Without this assumption the transition probabilities given in equation 24 are not technically correct.

The transition probabilities $\tau$, coupled with a policy based on the belief state defines a Markov process on the belief states. By adapting the reward function $R$ of the POMDP to a reward function $r$ on the belief states we have $r : A \times B \rightarrow \mathcal{R}^+$, the expected reward from taking action $a$ when in belief state $b$ is

$$
r(h, b) = q_h(1 - P_c/2) + E[q'](1 - (1 - P_c/2))
$$

(25)

$$
r(o, b) = \hat{q}_o(1 - P_c/2)(1 - P_c')^t + E[q'](1 - (1 - P_c/2)(1 - P_c')^t)
$$

(26)
Our goal is to find a policy \( \pi : \mathcal{B} \rightarrow \mathcal{A} \) which maximizes the expectation of
\[
\sum_{t=1}^{T} r(\pi(b_t), b_t)
\] (27)
where \( T \) is a geometrically distributed random variable, with parameter \( P_d \), so that the expectation of \( T \) is \( 1/P_d \). Thus the original POMDP has been transformed into a MDP on the belief states. Since the true state of the other patch \( q_o \) is never known by the forager, the \( \hat{v} \) of \( \hat{q}_o \) is dropped when discussing the belief MDP.

Before computing an approximation of the optimal policy it will be helpful to define the following value functions.

**Definition 3.1.** \( V(q_h, q_o, t) \) is the expected value of being in belief state \( b = (q_h, q_o, t) \), given that an optimal policy is employed. It can be defined recursively as

\[
V(q_h, q_o, t) = \max[V_h(q_h, q_o, t), V_o(q_h, q_o, t)]
\] (28)

where

\[
V_h(q_h, q_o, t) = (1 - P_d)((1 - P_c/2)(q_h + V(q_h, q_o, t + 1)) + (P_c/2)E[q' + V(q', q_o, t)])
\] (29)

\[
V_o(q_h, q_o, t) = (1 - P_d)((1 - P_c/2)(1 - P'_c)^t)V_{static}(q_h, q_o) + (1 - (1 - P_c/2)(1 - P'_c)^t)V_{change}(q_h))
\] (30)

with

\[
V_{static}(q_o, q_h) = q_o + V(q_o, q_h, 1)
\] (33)

and

\[
V_{change}(q_h) = E[q' + V(q', q_h, 1)]
\] (34)
Here $q'$ denotes a random variable drawn from the patch quality distribution (geometric with mean $1/\lambda$).

The standard approach for solving such problems is to let $\beta(t)$ be a vector representation of a probability distribution over the belief states $B$, and $\gamma_\pi$ be the transition matrix for the belief states, given that policy $\pi$ is employed, so that if $\beta(t)$ gives the probabilities of being in the various belief states at time $t$, then $\beta(t+1) = (\gamma_\pi \cdot \beta(t))$ gives the probability distribution over the belief states at time $t + 1$. Since $T$ is geometrically distributed the expected value of being in each state, when using policy $\pi$ is given by solving the following equation.

$$V_\pi(\beta) = (1 - P_d)(r(\pi(\beta), \gamma_\pi \cdot \beta)) + V_\pi(\gamma_\pi \cdot \beta))$$  \hspace{1cm} (35)

Once the expected value of each state under a given policy is known, the policy can be strictly improved by iterating through all of the possible states and changing the policy so that the probability of transitioning to higher valued states is increased and the probability of transitioning to lower valued states is decreased. This updated policy can then be used to determine the value of each state when this new policy is used, and these new valuations of the states can be used to improve the policy yet again. The issue with this approach is that if the state space is infinite, as is the case here, it is impossible to iterate through all possible states, improving the policy state by state. This difficulty is overcome, when possible, by using a sensible finite approximation of the true state space. In this problem since the belief state is discrete and only three dimensions using a finite approximation of the true state space is a viable option.

In addition it is likely possible to adapt the results of Krishnamurthy and Djonin’s “Structured Threshold Policies for Dynamic Sensor Scheduling - A partially observed
Markov decision approach” [15], to show that attention can be restricted to threshold policies. In this paper the authors give a set of sufficient conditions for the existence of the threshold policies for POMDP’s. However only a finite state space is considered, where as in the problem described above the state space is infinite, hence the need for some minor adaptation.

In this problem however we will take a different approach and directly exploit the particular structure of this problem.

This problem is similar to one presented in [25] and to the simple two sensor scheduling example in [15]. In [25] a forager must choose, each round, between exploiting a patch with fixed quality and exploiting a patch that varies stochastically between a quality that is higher than and a quality that is lower than the fixed patch’s quality. The forager seeks to optimize it’s resource intake, but only knows the quality of the variable patch the last time it exploited the variable patch. In [15] a decision maker must choose at each time step between using either an accurate and expensive sensor, or an inaccurate and cheap sensor, to effectively track a moving target. When the target being tracked by the sensors is far from some critical location then the cost function is such that it is desirable to use the cheap sensor, but when the target is near the critical location it is better to use the expensive sensor. Sense neither sensor is perfectly accurate, the true state of the target (its position relative to the critical location) is not known with certainty, giving rise to a POMDP. While the details of the two sensor example differ from the foraging problem presented earlier in this section, in both problems there is a similar dilemma of how much effort to expend on gathering costly information as a function of the current degree of uncertainty.
3.1.2 The form of an optimal policy

The goal of this section is to restrict the space of policies that need to be considered, when looking for an optimal policy. To that end we will describe policies of a certain simple form and then show that a policy of this form can be arbitrarily close to an optimal policy.

Definition 3.2. A policy is of the standard form if it can be determined as follows.

\[
\pi(q_h, q_o, t) = \begin{cases} 
    h & \text{if } q_o < q_h \text{ and } q_c \leq q_h \\
    h & \text{if } q_o < q_h < q_c \text{ and } t < t_c(q_h, q_o) \\
    h & \text{if } q_c \leq q_h < q_o, \text{ and } t > t_s(q_h, q_o) \\
    o & \text{if } q_h \leq q_o, \text{ and } q_h < q_c \\
    o & \text{if } q_o < q_h < q_c \text{ and } t \geq t_c(q_h, q_o) \\
    o & \text{if } q_h < q_o, \text{ and } q_h \geq q_c \text{ and } t \leq t_s(q_h, q_o) \\
\end{cases}
\]

(36)

Where \( t_c : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+ \), \( t_s : \mathbb{Z}^+ \times \mathbb{Z}^+ \to \mathbb{Z}^+ \), and \( q_c \in \mathbb{Z}^+ \).

Note that in the case where \( q_o < q_h \) and \( q_c \leq q_h \), \( t_c(q_h, q_o) \) need not be defined.

Definition 3.3. Let \( q_c^* \) be the smallest value of \( q_h \) for which

\[
\lim_{t \to \infty} V_h(q_h, q_o, t) \geq V_{\text{change}}(q_h).
\]

\( q_c^* \) is the minimum value of \( q_h \) such that checking in on the other patch is no longer worthwhile, given that the other patch has changed.

Lemma 3.4. There is an optimal policy of the standard form with \( q_c = q_c^* \)

Proof. Case a: If \( q_h \leq q_o \) and \( q_h < q_c^* \) then \( \pi^*(q_h, q_o, t) = o \) for all \( t \) since \( V_o(q_h, q_o, t) \) is a convex combination of \( V_{\text{static}}(q_h, q_o) \) and \( V_{\text{change}}(q_h) \) both of which are guaranteed to be greater than or equal to \( V_h(q_h, q_o, t) \) in this case.
Case b: If \( q_o \leq q_h < q_c^* \) then there exists a \( t_c(q_h, q_o) \) (possibly zero, though a state with \( t = 0 \) is not possible \( V_h \) and \( V_o \) are still defined for \( t = 0 \)) such that \( \pi^*(q_h, q_o, t) = h \) when \( t \leq t_c(q_h, q_o) \) and \( \pi^*(q_h, q_o, t) = o \) when \( t > t_c(q_h, q_o) \). \( V_o(q_h, q_o, t) \) is a convex combination of \( V_{\text{static}}(q_h, q_o) \) and \( V_{\text{change}}(q_h) \), with the weight given to \( V_{\text{change}} \) monotonically increasing with \( t \). Thus \( V_o(q_h, q_o, t) \) approaches \( V_{\text{change}}(q_h) \) in the limit as \( t \to \infty \). In contrast, since \( q_h < q_c^* \) and by the definition of \( q_c^* \), \( V_h(q_h, q_o, t) \) also monotonically increases but must approach some value strictly less than \( V_{\text{change}}(q_h) \). The desired result follows.

Case c: If \( c^* < q_h \leq q_o \) then there exists a \( t_s \) (possibly zero) such that \( \pi^*(q_h, q_o, t) = o \) when \( t \leq t_s \) and \( \pi^*(q_h, q_o, t) = h \) when \( t > t_s \). This follows from an analogous argument to that used in Case b.

Case d: If \( q_h > q_o \) and \( q_h > c^* \) then \( \pi^*(q_h, q_o, t) = h \) for all \( t \). This follows from an analogous argument to that used in Case a.

Thus an optimal policy can be described by the constant \( c^* \), and functions \( t_c^* \) and \( t_s^* \) and the problem of finding an optimal policy can be recast as finding the constant \( q_c^* \), and the functions \( t_c^* \) and \( t_s^* \). Further inspection suggests the following lemma.

**Lemma 3.5.** If an optimal policy is employed and the initial belief state does not satisfy \( q_c^* < q_h < q_o \), then no belief state will ever satisfy \( q_c^* < q_h < q_o \).

**Proof.** Suppose that a forager is in a belief state with \( q_c^* < q_h < q_o \). This can only have occurred if the forager was previously in a state where \( q_c^* < q_h < q_o \) and \( t > t_c^* \) or a state where \( q_c^* < q_o < q_h' \) and where \( q_h' \) has just recently changed to \( q_h \). This former type of state \( (q_c^* < q_h < q_o \text{ and } t > t_c^*) \) can only have occurred if the latter
type of state \((q^*_c < q_o < q_h)\) has occurred at some previous point. Thus to be in a state where \(q^*_c < q_h < q_o\) the forager must have previously been in a state where \(q^*_c < q_o < q_h\). The forager can only know that \(q^*_c < q_o\) if it has been at the patch with quality \(q_o\), and then switched to the patch with quality \(q_h\). Thus in order to be in a state of the form \(q^*_c < q_h < q_o\) a forager must have previously been in a state where \(q^*_c < q'_h < q_o\).

**Definition 3.6.** A policy is of the simple form if it can be determined as follows

\[
\pi(q_h, q_o, t) = \begin{cases} 
    h & \text{if } q_o < q_h \text{ and } t < t_c(q_h, q_o) \\
    o & \text{if } q_o < q_h \text{ and } t \geq t_c(q_h, q_o) \\
    o & \text{if } q_h \leq q_o
\end{cases}
\]  

(37)

**Proposition 3.7.** There is a policy of the simple form with performance arbitrarily close to the performance of an optimal policy provided the initial belief state of the forager is not of the form \(q^*_c < q_h < q_o\).

**Proof.** By lemma 3.4 there is an optimal policy of the standard form. Since the initial belief state does not satisfy \(q^*_c < q_h < q_o\) then by lemma 3.5 there will never be belief states where \(q^*_c < q_h < q_o\) and so an optimal policy is not affected in any way by the function \(t_s^*(q_h, q_o)\). That is the case where \(q_h < q_o\) and \(q_h \geq q_c\) and the case where \(q_c \leq q_h < q_o\) and \(t > t_s(q_h, q_o)\) need not be considered.

For all other cases except where \(q_o < q_h\) and \(q^*_c \leq q_h\) a policy of the simple form, using the optimal \(t^*_c(q_h, q_o)\), is identical to an optimal policy of the standard form.

In the \(q_o < q_h\) and \(q^*_c \leq q_h\) case an optimal policy will always employ the action \(h\) where as a policy of the simple form, since it uses the rule take action \(h\) if \(t < t_c(q_h, q_o)\) and take action \(o\) if \(t \geq t_c(q_h, q_o)\), will sometimes take action \(o\). The difference in expected value resulting from using two distinct policies is bounded by the probability
of those two policies producing different responses to the same situations. The probability that a policy of the simple form will take an action different from that of the optimal policy, in these cases is $\gamma^t_{c}(q_h, q_o)$ where $\gamma = (1 - P_d)(1 - P_c/2)$. Clearly then, for the case $q_o < q_h$ and $q^*_c \leq q_h$, as $t_c(q_h, q_o)$ approaches infinity an optimal simple policy converges on an optimal policy of the standard form.

This gives a convenient representation for near optimal policies, in that they can be totally determined by an integer function $t_c(q_h, q_o)$.

### 3.1.3 Approximating the optimal policy

It is clear that for $q_h \leq q_o$, $t^*_c(q_h, q_o) = 1$. It remains to compute an approximation of the optimal $t_c(q_h, q_o)$ for the case where $q_h > q_o$. Consider now that we are only interested in what the optimal value of $t_c(q_h, q_o)$ is and so we only need to know if the value of being in state $(q_h, q_o, 1)$ increases or decreases when $t_c(q_h, q_o)$ is incremented. Let $D = V_{t_c}(q_h, q_o, 1) - V_{t_c-1}(q_h, q_o, 1)$, $D$ for difference. If $D$ is positive, the performance of the policy will be improved by incrementing $t_c(q_h, q_o)$. Assuming that $V_{t_c}(q_h, q_o, 1)$ is concave in $t_c$, an approximation of the optimal $t_c$ can be found by incrementing $t_c$ until $D$ is negative. Note that if $q_h > q^*_c$ then $D$ will never be negative, and so for some cutoff $C$, we must assume that if $D > 0$ for all $t_c(q_h, q_o) \leq C$, then $q_h > q^*_c$, and a forager should never check in on the other patch.

We begin by defining some important quantities.

**Definition 3.8.** Let $r_1$ be the first time a change occurs, since the forager has transitioned to belief state $(q_h, q_o, 1)$. This means that a forager has just switched from a patch with quality $q_o$ to a patch with quality $q_h$. If we let $r_h$ be the first time the $h$...
patch changes, let \( r_o \) be the first time that the \( q_o \) quality patch changes, and let \( r_d \) be the time that the foraging session ends, we have that \( r_1 = \min[r_h, r_o, r_d] \).

Given the dynamics of the system, both \( r_h \) and \( r_o \) are geometrically distributed with parameter \( P_c/2 \). Note though that \( r_h \) and \( r_o \) are not independent. \( r_d \) is geometrically distributed with parameter \( P_d \). Let \( \delta = (1 - P_c)(1 - P_d) \) denote the probability that the session doesn’t end and neither patch changes quality on a given round. Thus \( r_1 \) is geometrically distributed so that \( \Pr(r_1 = k) = \delta^{k-1}(1 - \delta) \) with \( k \in [1, 2, 3, \ldots] \). Note that if \( r_1 = 1 \) then either the foraging session has ended or one of the patches has changed quality on the very round that the forager has switched from a \( q_o \) quality patch to a \( q_h \) quality patch.

We start by considering the contribution to \( D \) accrued before the first change occurs. If a forager checks in on the other patch after \( t_c \) rounds instead of after \( t_c - 1 \) rounds, this means that they will spend less time checking in on the \( q_o \) quality patch prior to any changes, and so there will be a slightly lower cost of checking. The number of times that a forager checks in on the lower quality patch prior to the first change occurring is

\[
\left\lfloor \frac{r_1 - 1}{t_c + 1} \right\rfloor
\]

where \( \lfloor \cdot \rfloor \) is the floor (or round down to the nearest integer) function. This number is distributed geometrically on \([0, 1, \ldots]\) with parameter \( 1 - \delta^{t_c+1} \). Conditioning on \( r_1 \) the expected difference in accrued value prior to the first change, resulting from checking less frequently is as follows.

\[
(q_o - q_h) \left( \frac{\delta^{t_c+1}}{1 - \delta^{t_c+1}} - \frac{\delta^{t_c}}{1 - \delta^{t_c}} \right)
= (q_h - q_o) \left( \frac{\delta^{t_c}(1 - \delta)}{(1 - \delta^{t_c+1})(1 - \delta^{t_c})} \right)
\]
Now that we have the contribution to $D$ accrued prior to the first change, we next consider the contributions to $D$ accrued during and after the first change. To compute these precisely requires either knowing the value function $V$ or equivalently conditioning on the type and timing of all possible subsequent changes (patch quality change or foraging session ending) to occur after the forager enters state $(q_h, q_o, 1)$. Since the value function $V$ is not known, nor is it feasible to condition on all possible subsequent changes, we can only approximate these contributions to $D$. This approximation is made by conditioning only on the type and timing of the first two changes. This is a straightforward, but long and tedious process and can be found in appendix A.2.

A program for computing the approximately optimal $t_c(q_h, q_o)$ by conditioning on the time and type of the first and second patch quality changes, as described in the appendix A.2, can be found in appendix B.1. Figure 1 shows a contour plot of this approximation of $t^*_c(q_h, q_o)$, for parameters $P_c = 0.1$, $P_d = 0.02$, and $\lambda = 0.1$.

### 3.2 Many Foragers, Free Switching, Variable Terminal Time, Dynamic Environment

We now consider the same problem as in the previous section except that now there is a large population of foragers exploiting the two patches instead of a single forager, and the dynamics of the patch qualities are slightly altered for analytic tractability.

Foragers would always like to be exploiting at the patch with the highest payoff per forager, but this is challenging for two reasons. The first is that, as in the lone forager problem, foragers only experience the quality of the patch that they exploit
at on a given round. The second is that because all of the foragers are trying to be at the patch with the highest payoff per forager, whenever there is a difference in expected payoff per forager at each patch there is an entire population competing to take advantage of this difference.

![Diagram showing the equilibrium proportion of foragers at the h patch as a function of q_o for the q_h values of 1, 5, 10, 20, 40, and 80.]

Suppose that a forager finds itself at a patch of quality 5, the second curve from the bottom in figure 2, and the other patch is of quality 20, and that the population is distributed in such a way that no forager has a preference for one patch over the other. When the population is distributed in this way we will say that it is in equilibrium. Because moving from one patch to the other has no cost, when the population is in
an equilibrium distribution the expected fitness of foragers at either patch must be equal. According to figure 2 this equilibrium distribution is when 0.21 of the foragers are at the quality 5 patch and 0.79 of the foragers are at the quality 20 patch. Note that if the environment were static the equilibrium distribution would have 0.2 of the foragers at the quality 5 patch, and 0.8 of the foragers at the quality 20 patch. The reason for this discrepancy is that in a changing environment, patch quality changes will create situations where the population distribution is out of equilibrium. The foragers at the patch with the minority of the population, are more likely to benefit from a patch quality change than foragers at the patch with the majority of the population, given these particular patch quality dynamics.

3.2.1 POMDP formulation

The standard POMDP formulation can be augmented with the tools of evolutionary game theory to precisely describe this problem. $h$ (for here) and $o$ (for other) are used to denote the patch that a forager has most recently foraged at, and the other patch respectively, relative to a specific forager, in the context where there is a specific forager, and are simply objective labels when there is no specific forager. Elements of the state space $S$ consist of the current quality of the two patches, $q_h$ and $q_o$, the proportion of the population at each patch $d_h$ and $d_o = 1 - d_h$, and a variable $I_c$ indicating the number of rounds until a patch quality change will be possible. Thus $s \in S$ is of the form $s = (q_h, q_o, d_h, I_c)$. The action space is $A = \{0, 1\}$, where $a = 1$ indicates a forager exploiting the same patch as on the previous round and $a = 0$ indicates a forager exploiting at the patch that it did not exploit at previously.

The quality aspect of the state transitions, $T((q_h^t, q_o^t), (q_h^{t+1}, q_o^{t+1}))$, are the same.
Problem 3.2 Summary

- Each round there is a patch quality change with probability $P_c$, except on the two rounds immediately following a patch quality change, in which case the probability of a patch quality change occurring is 0. When a patch changes quality each patch is equally likely to be the patch that changes.
- If a patch quality changes, its new value is drawn from a geometric distribution with parameter $\lambda$.
- Each round, with probability $P_d$, the foraging session ends.
- There is no explicit cost to switching from one patch to the other, only implicit opportunity costs.
- The population is large enough that the proportion of the population employing a given action on a round is effectively deterministic, even if the action employed by any one individual is the stochastic result of a mixed strategy.
- A forager knows
  - The quality of its current patch.
  - The proportion of the population at each patch.
  - The quality of the patch it is not currently at, the last time it was at that patch.
  - The parameters of the model, $P_c$, $P_d$ and $\lambda$.
  - The policy used by the population, so that inferences about the quality of the patch it is not currently at can be made.
- A forager does not know
  - The current quality of the patch it is not currently at.

What we are looking for: The equilibrium distribution of foragers at each patch and the probability of switching from one patch to the other at these patches such that this equilibrium distribution is maintained in the absence of patch quality changes. Both switching probabilities and the equilibrium distribution are dependent on the quality of both patches.
as in the previous lone forager problem, except that to make the model more readily tractable, there is now a “cool down” period between detectable patch quality changes. The effect of this is that if a patch changes quality on round $t$, say, then on round $t + 1$ and $t + 2$ there is no possibility of a patch quality change. Every round after $t + 2$ there will be a patch quality change with probability $P_c$ until there is another change which will be followed by another cool down period. Thus the number of rounds between patch quality changes is a geometric random variable shifted up by two. The quality transition probabilities are as follows

$$T((q_h, q_o, I_c), (q'_h, q'_o, 0)) = 1 - P_c + (P_c/2)\lambda(1 - \lambda)^{q_h-1} + (P_c/2)\lambda(1 - \lambda)^{q_o-1}$$

if $q_h = q'_h$ and $q_o = q'_o$ and $I_c = 0$ (40)

$$T((q_h, q_o, I_c), (q'_h, q'_o, I_c - 1)) = 1$$

if $q_h = q'_h$ and $q_o = q'_o$ and $I_c \geq 1$ (41)

$$T((q_h, q_o, I_c), (q'_h, q'_o, 2)) = (P_c/2)\lambda(1 - \lambda)^{q'_h-1}$$

if $q_h \neq q'_h$ and $q_o = q'_o$ and $I_c = 0$ (42)

$$T((q_h, q_o, I_c), (q'_h, q'_o, 2)) = (P_c/2)\lambda(1 - \lambda)^{q'_o-1}$$

if $q_h = q'_h$ and $q_o \neq q'_o$ and $I_c = 0$ (43)

Note that the patch quality transitions are totally independent of the actions of the foragers. In contrast the population distribution aspect of the state is totally determined by the cumulative actions of the population.

$$d_{h}^{t+1} = E[a|forager \ at \ h \ patch](d_h^t) + (1 - E[a|forager \ at \ o \ patch])(1 - d_h^t)$$ (44)

Note the implicit assumption here that the population is large enough that the dynamics of the population distribution can be treated as deterministic with respect to
expectations of the proportion of foragers at a given patch that will take a given action, despite the fact that an individual forager’s action may be stochastically chosen.

The reward function for a forager employing action $a$ in state $s^t = (q^t_h, q^t_o, d^t_h, I^t_c)$ then transitioning to state $s^{t+1} = (q^{t+1}_h, q^{t+1}_o, d^{t+1}_h, I^{t+1}_c)$ is

$$R_a(s^t, s^{t+1}) = \frac{q^{t+1}_h}{d^{t+1}_h}$$

The objective of an individual forager is to optimize the following

$$\sum_{t=1}^{T} R_a(s^t, s^{t+1})$$

where $T$ is a geometrically distributed random variable with parameter $P_d$. A forager does this by adopting a policy function, $\pi : B \to [0, 1]$, which gives the probability with which a forager should exploit at the same patch on the next round as it did in the current round, given the current belief state $b \in B$ of the forager.

The population is assumed to be large enough that the policy of a single forager has a negligible impact on the dynamics of the population distribution aspect of the state. In general for any population employing policy $\pi_r$ ($r$ for resident) there will be at least one policy $\pi_{B(r)}$ ($B$ for best response) which optimizes expression 46, given that effectively all of the population is employing policy $\pi_r$. We are looking for a policy $\pi^*$ with the interesting property that if the entire population adopts policy $\pi^*$ then no policy other than $\pi^*$ will be the optimizing policy for the state dynamics that a population using $\pi^*$ creates, that is an evolutionarily stable (ES) policy.

Each time step a forager observes the value of the patch it has just foraged at, the proportion of the population that also foraged at that same patch, and whether the patch that it has just foraged at is the patch that it foraged at in the previous time step or the other patch. Thus the observation set is $O = A \times Z^+ \times [0, 1]$. Since there is no noise on this observation, the new information a forager receives when they take action
a and transition from state $s^t = (q^t_h, q^t_o, d^t_h, I^t_c)$ to state $s^{t+1} = (q^{t+1}_h, q^{t+1}_o, d^{t+1}_h, I^{t+1}_c)$ is $o^{t+1} = (a^t, q^{t+1}_h, d^{t+1}_h)$ with probability one. The information available to the forager is its entire history of observations.

Assuming that the foragers have full knowledge of the dynamics of the patch qualities and that the foragers have full knowledge of the policy $\pi$ employed by the population, then depending on the particular policy employed by the population, a forager may or may not be able to infer information about the quality of the patch that it has not just foraged at, that is $q^{t-1}_o$. Thus not only the dynamics of a forager’s belief state but also the very information that might inform this belief state is dependent on the policy adopted by the population. Conversely a policy cannot be defined without reference to the structure of the information available to the foragers. To overcome this difficulty we will first define a specific form of policy without worrying whether or not it is based on information that a forager actually has access to. Then we will show that if this form of policy is employed by the population, individual foragers will indeed have access to the information required by the policy.

### 3.2.2 The form of policies

Before characterizing a general form of policy the following definitions are needed.

**Definition 3.9.** Let $V_\pi(b)$ denote the total expected reward a forager receives given that its initial belief state is $b$ and that it uses policy $\pi$ given that the population is also employing policy $\pi$.

**Definition 3.10.** The state of the system is said to be in equilibrium, with respect to a policy $\pi$ employed by the population, if all of the following hold.
• The belief states of the foragers are completely determined by their location, that
is all the foragers at a patch all share the same belief state $b$ and the foragers at
the other patch all have belief state $b'$.

• The expected value resulting from being at either patch is equal, that is

$$V_{\pi}(b) = V_{\pi}(b')$$  (47)

• The proportion of foragers that leave one patch will be perfectly replaced by
foragers arriving from the other patch. That is

$$(1 - \pi(b))(d) = (1 - \pi(b'))(1 - d)$$  (48)

where $d$ is the proportion of the population at the patch where the foragers all
have belief state $b$

• The state is such that a quality change was possible, though did not occur, on
this current round, which is denoted, in a slight stretch of notation, by $I_c = -1$.

**Definition 3.11.** An equilibrium population distribution function $d^\star(q_h, q_o)$, gives the
proportion of the population at the $h$ patch when the system is in equilibrium and the
patch qualities are $q_h$ and $q_o$.

**Definition 3.12.** The transitional population distribution function $d^\#(q_h, q_o)$ gives the
proportion of the population at the $h$ patch such that the reward per forager at each
patch is equal given that the other patch has quality $q_o$.

$$d^\#(q_h, q_o) = \frac{q_h}{q_h + q_o}$$  (49)

**Definition 3.13.** An equilibrium switching regime $s^\star(q_h, q_o) = s^\star(q_o, q_h)$ is the propor-
tion of the total population switching from the $h$ patch to the $o$ patch, or equivalently
the proportion of the total population switching from the o patch to the h patch when
the state is in equilibrium.

A policy can be primarily described in terms of a target population distribution
function $d^*: \mathbb{Z}^+ \times \mathbb{Z}^+ \to [0, 1]$, a switching function $s^*: \mathbb{Z}^+ \times \mathbb{Z}^+ \to [0, 1]$, and the
transitional population distribution function $d^t: \mathbb{Z}^+ \times \mathbb{Z}^+ \to [0, 1]$.

A policy $\pi$ is said to be of the standard form if it satisfies the following.

Case A. In the case where a forager knows $q_{t-1}^o$ with certainty, the forager’s patch
has not changed, $q_{t}^h = q_{t-1}^h$, a change could have occurred last round, and
the current population distribution is the target population distribution of the
policy, then the policy is

$$\pi(q_t^h, q_{t-1}^h, d^*(q_t^h, q_{t-1}^o), d^{t-1}, q_{t-1}^o, -1) = 1 - \frac{s^*(q_{t-1}^h, q_{t-1}^o)}{d^t}$$ (50)

The first and second arguments of the policy function, here taking values $q_{t}^h$ and
$q_{t-1}^h$, are the quality of the patch the forager is currently at, on the current
and previous round respectively. The third and fourth arguments of the policy
function, here $d^*(q_t^h, q_{t-1}^o)$ and $d^{t-1}$, refer to the proportion of the population
that exploited at the focal forager’s current patch on the current round and the
previous round respectively. The fifth argument of the policy function, here
$q_{t-1}^o$, is the quality of the patch that the forager did not exploit at this round,
on the previous round. The sixth argument of the policy function, here $-1$,
indicates the number of rounds until a patch quality change is possible. In a
slight stretch of notation the $-1$ indicates that there may have been a patch
quality change at the other patch this very round, but if there has not, then
there may be a patch quality change at either patch on the next round.
Case B. In the case where a forager knows both $q_{t-1}^h$ and $q_{t-1}^o$ with certainty, the forager’s patch has just changed, $q_t^h \neq q_{t-1}^h$, and the current population distribution would have been the target population distribution had the forager’s patch not changed quality, then the policy is

$$\pi(q_t^h, q_{t-1}^h, d^*(q_{t-1}^h, q_{t-1}^o), d_{t-1}, q_{t-1}^o, 2) = \begin{cases} \frac{d^2(q_t^h, q_{t-1}^o) - s^*(q_{t-1}^h, q_{t-1}^o)}{d^2} & \text{if } 0 < \frac{d^2(q_t^h, q_{t-1}^o) - s^*(q_{t-1}^h, q_{t-1}^o)}{d^2} < 1 \\ 0 & \text{if } \frac{d^2(q_t^h, q_{t-1}^o) - s^*(q_{t-1}^h, q_{t-1}^o)}{d^2} \leq 0 \\ 1 & \text{if } 1 \leq \frac{d^2(q_t^h, q_{t-1}^o) - s^*(q_{t-1}^h, q_{t-1}^o)}{d^2} \end{cases}$$

(51)

The 2 in the last argument of the policy function is to indicate that the forager knows that there will not be another patch quality in the next two rounds, that is $I_c = 2$.

Case C. In the case where a forager knows both $q_{t-1}^h$ and $q_{t-1}^o$ with certainty, and further that a change was not possible this round and will not be possible on the next round and the current population distribution is the transitional population distribution for the current patch qualities, then the policy is

$$\pi(q_t^h, q_{t-1}^h, d^*(q_{t-1}^h, q_{t-1}^o), d_{t-1}, q_{t-1}^o, 1) = 1$$

(52)

Since the population is distributed in such a way that the per round reward per forager is equal at both patches, and a patch quality change is not possible, none of the foragers switch patches. The 1 in the last argument indicates that
the forager knows that there was a change on the previous round and so there can not have been a quality change on this round, nor can there be a change in quality on the next round.

Case D. In the case where a forager knows both \( q_{t-1}^h \) and \( q_{t-1}^o \) with certainty, and further that a change was not possible on this round but will be possible on the next round, and the current population distribution satisfies \( d_t < d^* (q_{t-1}^h, q_{t-1}^o) \), then the policy is

\[
\pi(q_t^h, q_{t-1}^h, d_t, d_{t-1}, q_{t-1}^o, 0) = 1 \tag{53}
\]

Case E. In the case where the forager knows both \( q_{t-1}^h \) and \( q_{t-1}^o \) with certainty, and further that a change was not possible on this round but will be possible on the next round, and the current population distribution satisfies \( d_t > d^* (q_{t-1}^h, q_{t-1}^o) \), then the policy is

\[
\pi(q_t^h, q_{t-1}^h, d_t, d_{t-1}, q_{t-1}^o, 0) = \frac{d^*(q_t^h, q_{t-2}^o)}{d_t} \tag{54}
\]

Case F. In the case where a forager does not know the precise quality of the other patch, only that the quality of the other patch has changed to something worse than some upper bound on the previous round, so that there were, and now are, “too many” foragers at the other patch, that is \( d_t < d^*(q_{t-1}^h, q_{t-1}^o) \), the policy is

\[
\pi(q_t^h, q_{t-1}^h, d_t, d_{t-1}, -2, 1) = 1 \tag{55}
\]

Where the \(-2\) in the fifth argument of the policy function indicates that the forager does not know \( q_{t-1}^o \) precisely, and can only infer an upper bound on its value. Since a forager in this case is ignorant of the quality of the other patch, it leaves it to the informed foragers at the patch with too many foragers to establish the transitional target population distribution.
Case G. In the case where a forager has just left a patch that drastically worsened on the previous round the policy is

$$\pi(q^t, q^{t-1}_h, d^t, d^{t-1}, q^{t-1}_o, 1) = 1 \quad (56)$$

Like the ignorant foragers this forager shares a patch with, a forager in this case leaves it to the informed foragers at the patch with too many foragers, to establish the transitional target population distribution.

Case H. In the case where a forager arrives at a patch and finds that its quality has worsened on the previous round and as a result there are “too many” foragers at its current patch, that is $d^t > d^\sharp(q^t_h, q^{t-1}_o)$, the policy is

$$\pi(q^t, q^{t-1}_h, d^t, d^{t-1}, q^{t-1}_o, 1) = \frac{d^\sharp(q^t_h, q^{t-1}_o)}{d^t} \quad (57)$$

Case I. In the case where a forager does not know the precise quality of the other patch, only that the quality of the other patch has changed to something better than some lower bound on the previous round, so that there were, and now are, “too few” foragers at the other patch, that is $1 - d^t < d^\sharp(q^{t-1}_o, q^{t-1}_h)$, the policy is

$$\pi(q^t, q^{t-1}_h, d^t, d^{t-1}, -1, 1) = 0 \quad (58)$$

The $-1$ in the fifth argument of the policy function indicates that the forager does not know $q^{t-1}_o$ precisely, and can only infer a lower bound on its value.

Case J. In the case where a forager has either just arrived at some drastically improved patch, or has been at that patch since it improved, $d^t < d^\sharp(q^t_h, q^{t-1}_o)$ and
so knows both $q_{o}^{t-1}$ and $q_{h}^{t-1}$ with certainty, the policy is

$$
\pi(q_{h}^{t}, q_{h}^{t-1}, d^t, d^{t-1}, q_{o}^{t-1}, 1) = \begin{cases} 
\frac{d^t(q_{h}^{t}, q_{o}^{t-1}) - (1-d^t)}{d^t} & \text{if } 0 < \frac{d^t(q_{h}^{t}, q_{o}^{t-1}) - (1-d^t)}{d^t} \\
0 & \text{if } \frac{d^t(q_{h}^{t}, q_{o}^{t-1}) - (1-d^t)}{d^t} \leq 0 
\end{cases} 
$$

(59)

The foragers at this drastically improved patch anticipate that all of the foragers from the other patch will be switching to the improved patch, and so choose a policy that will establish the transitional target population distribution, if possible.

**Proposition 3.14.** When a policy of the standard form is employed by the population as a whole, and all of the foragers know this, then the information required by a policy of the standard form can be inferred from a forager’s history, assuming that the forager knows the policy employed by the population and the patch quality dynamics.

**Proof.** Policies of the standard form take as their input, belief states of the form $b = (q_{h}^{t}, q_{h}^{t-1}, d^t, d^{t-1}, q_{o}^{t-1}, I_c)$, with $q_{h}^{t}, q_{h}^{t-1} \in \mathbb{Z}^+$, $d^t, d^{t-1} \in (0, 1)$, $q_{o}^{t-1} \in \mathbb{Z}^+ \cup \{-1, -2\}$, and $I_c \in \{-1, 0, 1, 2\}$.

$q_{h}^{t}, d^t$, and $d^{t-1}$ are all available to a forager directly from its history.

In the case where a forager stays at the same patch then $q_{h}^{t-1}$ is also directly available from the history, but $q_{o}^{t-1}$ and $I_c$ must be inferred from the history using knowledge of the policy employed by the population and knowledge of the patch dynamics.

A forager makes these inferences as follows. Suppose a forager is in an equilibrium belief state, that is a belief state of the form $(q_{h}, q_{o}, d^*(q_{h}, q_{o}), d^{t-1}, q_{o}, -1)$. If on the next round $d^t = d^*(q_{h}, q_{o})$ the forager can infer that the other patch did not change quality on the previous round, and that $q_{o}^{t-1} = q_{o}$ and that $I_c = -1$. If $d^t = 1 - s^*(q_{h}, q_{o})$ the forager can infer that the other patch has changed to something
worse than some lower bound on the previous round, and so \( q_o^{t-1} = -2 \) and \( I_c = 1 \). If \( d^t = d^*(q_h, q_o) - s^*(q_h, q_o) \) the forager can infer that the other patch has changed to something greater than some upper bound on the previous round, and so \( q_o^{t-1} = -1 \) and \( I_c = 1 \). If \( d^*(q_h, q_o) - s^*(q_h, q_o) < d^t < 1 - s^*(q_h, q_o) \) and \( d^t \neq d^*(q_h, q_o) \) the forager can infer that the other patch has changed on the previous round to a quality such that \( d^t = d^*(q_h^t, q_o^{t-1}) \) holds and that \( I_c = 1 \). In the non-equilibrium belief states, since there is a minimum of two rounds between patch quality changes, a forager knows that \( q_o^t = q_o^{t-1} \), and so can always infer the quality of the other patch on the previous round and \( I_c \).

When a forager has switched patches, the reverse is true, so that \( q_o^{t-1} \) is directly available from the history, and \( q_h^{t-1} \) must be inferred using knowledge of the policy and the quality dynamics. These inferences are made in the same way as in the case where a forager stays at a patch.

3.2.3 Dynamics of the Population’s Belief States

Rather than describing the dynamics of the belief states of an individual forager explicitly, it will be more informative to describe the dynamics of the set, \( P \), of population belief states profiles. \( P \) can be partitioned into distinct classes as follows. Before describing this partition of \( P \) the following quantities are needed.

**Definition 3.15.** Let \( q_b(q_h, q_o, d, s) \) denote the largest integer for which

\[
\frac{d^t(q_h, q_o) - s}{1 - d} \leq 0
\]

holds. \( q_b(q_h, q_o) \) can be computed as

\[
\left\lfloor \frac{q_h \cdot s}{1 - s} \right\rfloor
\]
where \( d = d^*(q_h, q_o) \) and \( s = s^*(q_h, q_o) \).

Thus if a patch of quality \( q_o \) changes to a patch of quality \( q'_o \leq q_b \), in a population of foragers using a policy of the standard form, all of the foragers at the patch that changes will switch to the \( q_h \) quality patch that did not change. When \( q_b \) appears its arguments will typically be suppressed as they will be unambiguous given the context.

**Definition 3.16.** Let \( q_B(q_h, q_o, d, s) \) denote the smallest integer for which

\[
1 \leq \frac{d^2(q_B, q_h) - s}{1 - d}
\]

holds. \( q_B(q_h, q_o) \) can be computed as

\[
\left\lfloor \frac{q_h(1 - d + s)}{1 - (1 - d + s)} \right\rfloor
\]

where \( d = d^*(q_h, q_o) \) and \( s = s^*(q_h, q_o) \).

Thus if a patch of quality \( q_o \) changes to a patch of quality \( q'_o \geq q_B \), in a population of foragers using a policy of the standard form, all of the foragers at the patch that just changed will stay at that patch. When \( q_B \) appears its arguments will typically be suppressed as they will be unambiguous given the context.

**Definition 3.17.** Let \( q_{BB}(q_h, q_o) \) denote the smallest integer for which

\[
d - s \leq d^2(q_{BB}, q_h)
\]

holds. \( q_{BB}(q_h, q_o) \) can be computed as

\[
\left\lfloor \frac{q_h(d - s)}{1 - (d - s)} \right\rfloor
\]

where \( d = d^*(q_h, q_o) \) and \( s = s^*(q_h, q_o) \).

Thus if a patch of quality \( q_o \) changes to a patch of quality \( q_B \leq q'_o \) then one round after that change occurs there will be \( d^*(q_h, q_o) - s^*(q_h, q_o) \) foragers at the \( q_h \) quality.
patch, and \( d^*(q_o, q_h) + s^*(q_h, q_o) \) at the \( q_o' \) quality patch, since all of the foragers at the improved patch stay where they are on the first round after the change. If \( q_o' < q_{BB} \) then two rounds after this change occurs there will be \( d^*(q_h, q_o) - s^*(q_h, q_o) \) foragers at the \( q_o' \) quality patch, and \( 1 - d^*(q_h, q_o) + s^*(q_h, q_o) \) foragers at the \( q_h \) quality patch, since all of the foragers at the unimproved patch switch to the improved patch, and all of the foragers at the improved patch switch to the unimproved patch. The population is then in a situation where there are too few foragers at one patch and too many foragers at the other, and the foragers at the patch with too few foragers will leave it to the foragers at the patch with too many foragers to reestablish equilibrium. If on the other hand \( q_{BB} \leq q_o' \) then two rounds after this change occurs the population will be in the appropriate transitional population distribution. This is assuming that the population is using a policy of the standard form. When \( q_{BB} \) appears its arguments will typically be suppressed as they will be unambiguous given the context.

Let \( E \) (for equilibrium) denote the subset of \( \mathcal{P} \), where the foragers at one patch have a belief states of the form

\[
b_h = (q_h^t, q_h^{t-1}, d^*(q_h^t, q_h^{t-1}), d^{t-1}, q_o^{t-1}, -1)
\]

and all of the foragers at the other patch have belief state

\[
b_o = (q_o^t, q_o^{t-1}, d^*(q_o^t, q_o^{t-1}), 1 - d^{t-1}, q_h^{t-1}, -1).
\]

In addition \( q_h^t = q_h^{t-1} \) and \( q_o^t = q_o^{t-1} \) both hold. Thus all of the foragers are in belief states of the form in Case A of the standard form policy description.

Let \( C \) (for change) denote the subsets of \( \mathcal{P} \), resulting from a change in patch quality in the case where the qualities of both patches on the previous round are known to all foragers. That is the previous belief state profile belonged to subset \( E \), \( W_0 \), \( W_- \), \( W_+ \) or \( W_{++} \) (\( W_0 \), \( W_- \), \( W_+ \) and \( W_{++} \) will be defined shortly). In these \( C \)
subsets of $P$ the foragers at one patch have a belief state of the form

$$b_h = (q^t_h, q^{t-1}_h, d^*(q^t_h, q^{t-1}_o), d^{t-1}, q^{t-1}_o, -1)$$

and the foragers at the other patch have a belief state of the form

$$b_o = (q^t_o, q^{t-1}_o, d^*(q^t_o, q^{t-1}_h), 1 - d^{t-1}, q^{t-1}_h, 2).$$

In addition $q^t_h = q^{t-1}_h$ and $q^t_o \neq q^{t-1}_o$, and $d^t = d^*(q^{t-1}_h, q^{t-1}_o)$ all hold. Thus the foragers at the patch that changed are in a belief state of the form in Case B of the policy description and the foragers at the patch that did not change are in a belief state of the form in Case A of the policy description. This subset of profiles is then further divided conditional on the type of change to occur.

Let $C_0$ denote the subset of $P$ where $q_b < q^t_o < q_B$.

Let $C_-$ denote the subset of $P$ where $q^t_o \leq q_b$.

Let $C_+$ denote the subset of $P$ where $q_B \leq q^t_o$ and $q_{BB} \leq q^t_o$.

Let $C_{++}$ denote the subset of $P$ where $q_B \leq q^t_o < q_{BB}$.

Let $R_0$ (for recovery) denote the subset of $P$ where all of the foragers at one patch have belief state

$$b_h = q^t_h, q^{t-1}_h, d^*(q^t_h, q^{t-1}_o), d^{t-1}, q^{t-1}_o, 1$$

and the foragers at the other patch have belief state

$$b_o = q^t_o, q^{t-1}_o, d^*(q^t_o, q^{t-1}_h), d^{t-2}, q^{t-1}_h, q^{t-1}_o, 1).$$

In addition $q^t_h = q^{t-1}_h = q^{t-2}_h$ and $q^t_o = q^{t-1}_o \neq q^{t-2}_o$ and $q_b < q^t_o < q_B$ all hold. Thus all of the foragers are in a belief state of the form in Case C of the policy description.

Let $R_-$ denote the subset of $P$ where $d^*(q^{t-2}_o, q^{t-1}_h)$ of the foragers at one patch have a belief state

$$b^1_h = (q^t_h, q^{t-1}_h, 1 - s^*(q^{t-1}_h, q^{t-2}_o), d^*(q^{t-1}_h, q^{t-2}_o), q^{t-1}_o, 1)$$
and the other $d^*(q_h^{t-1}, q_o^{t-2}) - s^*(q_h^{t-1}, q_o^{t-2})$ foragers at that same patch have belief state

$$b_h^2 = (q_h^t, q_h^{t-1}, 1 - s^*(q_h^{t-1}, q_o^{t-2}), d^*(q_h^{t-1}, q_o^{t-2}), -2, 1)$$

and the foragers at the other patch have belief state

$$b_o = (q_o^t, q_h^{t-1}, q_o^{t-2}, d^*(q_o^{t-2}, q_h^{t-1}), q_h^{t-1}, 1).$$

In addition $q_h^t = q_h^{t-1} = q_h^{t-2}$ and $q_o^t = q_o^{t-1} \neq q_o^{t-2}$ and $q_o^t \leq q_h$ all hold. Thus the foragers at the patch that changed are in a belief state of the form in Case H of the policy description, and the foragers at the patch that did not change are in belief states of the form in Case G and the form in Case F of the policy description.

Let $R_+$ denote the subset of $\mathcal{P}$ where the foragers at one patch have belief state

$$b_h = (q_h^t, q_h^{t-1}, d^*(q_h^{t-1}, q_o^{t-2}) - s^*(q_h^{t-1}, q_o^{t-2}), d^*(q_h^{t-1}, q_o^{t-2}), -1, 1)$$

and where all of the foragers at the other patch have belief state

$$b_o = (q_o^t, q_h^{t-1}, d^*(q_o^{t-2}, q_h^{t-1}) + s^*(q_h^{t-1}, q_o^{t-2}), d^*(q_o^{t-2}, q_h^{t-1}), q_h^{t-1}, 1).$$

In addition $q_h^t = q_h^{t-1} = q_h^{t-2}$ and $q_o^t = q_o^{t-1} \neq q_o^{t-2}$ and $q_B \leq q_o^t$ and $q_{BB} \leq q_o^t$ all hold. Thus the foragers at the patch that changed are in a belief state of the form in Case J of the policy description, and the foragers at the patch that did not change are in a belief state of the form in Case I of the policy description.

Let $R_{++}$ denote the subset of $\mathcal{P}$ where the foragers at one patch have belief state

$$b_h = (q_h^t, q_h^{t-1}, d^*(q_h^{t-1}, q_o^{t-2}) - s^*(q_h^{t-1}, q_o^{t-2}), d^*(q_h^{t-1}, q_o^{t-2}), -1, 1)$$

and where all of the foragers at the other patch have belief state

$$b_o = (q_o^t, q_o^{t-1}, d^*(q_o^{t-2}, q_h^{t-1}) + s^*(q_h^{t-1}, q_o^{t-2}), d^*(q_o^{t-2}, q_h^{t-1}), q_h^{t}, 1).$$
In addition \( q^t_h = q^{t-1}_h = q^{t-2}_h \) and \( q^t_o = q^{t-1}_o \neq q^{t-2}_o \) and \( q_B \leq q^t_o < q_{BB} \) all hold. Thus the foragers at the patch that changed are in a belief state of the form in Case J of the policy description, and the foragers at the patch that did not change are in a belief state of the form in Case I of the policy description.

Let \( W_0 \) (for waiting) denote the subset of \( \mathcal{P} \) where all of the foragers at one patch have belief state

\[
\begin{align*}
    b_h &= (q^t_h, q^{t-1}_h, d^*(q^t_h, q^{t-1}_o), d^*(q^{t-2}_h, q^{t-3}_o), q^{t-1}_o, 0) \\
    b_o &= (q^t_o, q^{t-1}_o, d^*(q^t_o, q^{t-1}_h), d^*(q^{t-2}_o, q^{t-3}_h), q^{t-1}_h, 0).
\end{align*}
\]

In addition \( q^t_h = q^{t-1}_h = q^{t-2}_h \) and \( q^t_o = q^{t-1}_o = q^{t-2}_o \neq q^{t-3}_o \) and \( q_h < q^t_i < q_B \) all hold. The foragers at one patch are in a belief state of the form in Case E of the policy description and the foragers at the other patch are in a belief state of the form in Case D of the policy description.

Let \( W_- \) denote the subset of \( \mathcal{P} \) where the foragers at one patch have belief state

\[
\begin{align*}
    b_h &= (q^t_h, q^{t-1}_h, d^*(q^t_h, q^{t-1}_o), 1 - s^*(q^{t-2}_h, q^{t-3}_o), q^{t-1}_o, 0) \\
    b_o &= (q^t_o, q^{t-1}_o, d^*(q^t_o, q^{t-1}_h), s^*(q^{t-2}_o, q^{t-3}_h), q^{t-1}_h, 0).
\end{align*}
\]

In addition \( q^t_h = q^{t-1}_h = q^{t-2}_h = q^{t-3}_h \) and \( q^t_o = q^{t-1}_o = q^{t-2}_o \neq q^{t-3}_o \) and \( q^t_o \leq q_h \) all hold. The foragers at one patch are in a belief state of the form in Case E of the policy description and the foragers at the other patch are in a belief state of the form in Case D of the policy description.

Let \( W_+ \) denote the subset of \( \mathcal{P} \) where the foragers at one patch have belief state

\[
\begin{align*}
    b_h &= (q^t_h, q^{t-1}_h, d^*(q^t_h, q^{t-1}_o), d^*(q^{t-2}_h, q^{t-3}_o) - s^*(q^{t-2}_h, q^{t-3}_o), q^{t-1}_o, 0) \\
\end{align*}
\]
and foragers at the other patch have belief state

\[ b_o = (q_o, q_o^{-1}, d^*(q_o, q_o^{-1}), d^*(q_o^{t-3}, q_h), s^*(q_h^{t-2}, q_o^{t-3}), q_h, 0). \]

In addition \( q_h^t = q_h^{t-1} = q_h^{t-2} = q_h^{t-3} \) and \( q_o^t = q_o^{t-1} = q_o^{t-2} \neq q_o^{t-3} \) and \( q_B \leq q_o \) and \( q_{BB} \leq q_o^t \) all hold. The foragers at one patch are in a belief state of the form in Case E of the policy description and the foragers at the other patch are in a belief state of the form in Case D of the policy description.

Let \( W_{++} \) denote the subset of \( \mathcal{P} \) where the the foragers at one patch have belief state

\[ b_h = (q_h^t, q_h^{t-1}, d^*(q_h^{t-3}, q_h^{t-2}), s^*(q_h^{t-2}, q_o^{t-3}), d^*(q_h^{t-3}, q_h^{t-2}) - s^*(q_h^{t-2}, q_o^{t-3}), q_h^{t-1}, 0) \]

and the foragers at the other patch have belief state

\[ b_o = q_o^t, q_o^{t-1}, d^*(q_h^{t-3}, q_h^{t-2}) - s^*(q_h^{t-2}, q_o^{t-3}), d^*(q_h^{t-3}, q_h^{t-2}) + s^*(q_h^{t-2}, q_o^{t-3}, q_h^{t-1}, 0). \]

In addition \( q_h^t = q_h^{t-1} = q_h^{t-2} = q_h^{t-3} \) and \( q_o^t = q_o^{t-1} = q_o^{t-2} \neq q_o^{t-3} \) and \( q_B \leq q_o^{t} < q_{BB} \) all hold. The foragers at one patch are in a belief state of the form in Case E of the policy description and the foragers at the other patch are in a belief state of the form in Case D of the policy description.

The following quantities will be used frequently in describing the dynamics of \( \mathcal{P} \).

\textbf{Definition 3.18.} Given that system is in a state with patch qualities \( q_h \) and \( q_o \) and
\( I_c = 0 \) then let

\[ \delta^0_h = P_c/2 \left( (1 - \lambda)^{q_h(q_o,q_h)} - (1 - \lambda)^{q_B(q_o,q_h)-1} \right) \]  
(66)

\[ \delta^0_o = P_c/2 \left( (1 - \lambda)^{q_o(q_o,q_o)} - (1 - \lambda)^{q_B(q_o,q_o)-1} \right) \]  
(67)

\[ \delta^0 = \delta^0_h + \delta^0_o \]  
(68)

\[ \delta^0_h = P_c/2 \left( 1 - (1 - \lambda)^{q_h(q_o,q_o)} \right) \]  
(69)

\[ \delta^0_o = P_c/2 \left( 1 - (1 - \lambda)^{q_o(q_o,q_o)} \right) \]  
(70)

\[ \delta^- = \delta^0_h + \delta^0_o \]  
(71)

\[ \delta^+_h = P_c/2 \left( (1 - I_{(q_{BB}>q_B)}) (1 - \lambda)^{q_h(q_o,q_h)-1} + I_{(q_{BB}>q_B)} (1 - \lambda)^{q_B(q_o,q_h)-1} \right) \]  
(72)

\[ \delta^+_o = P_c/2 \left( (1 - I_{(q_{BB}>q_B)}) (1 - \lambda)^{q_o(q_o,q_o)-1} - I_{(q_{BB}>q_B)} (1 - \lambda)^{q_B(q_o,q_o)-1} \right) \]  
(73)

\[ \delta^+ = \delta^+_h + \delta^+_o \]  
(74)

\[ \delta^{++}_h = P_c/2 \left( I_{(q_{BB}>q_B)} \right) \left( (1 - \lambda)^{q_h(q_o,q_h)-1} - (1 - \lambda)^{q_B(q_o,q_h)} \right) \]  
(75)

\[ \delta^{++}_o = P_c/2 \left( I_{(q_{BB}>q_B)} \right) \left( (1 - \lambda)^{q_o(q_o,q_o)-1} - (1 - \lambda)^{q_B(q_o,q_o)} \right) \]  
(76)

\[ \delta^{++} = \delta^{++}_h + \delta^{++}_o \]  
(77)

Where \( I_{(q_{BB}>q_B)} \) is an indicator function that takes value 1 when \( q_{BB} > q_B \) and the value 0 when \( q_{BB} \leq q_B \).

Then the transition probability matrix for the subsets of \( P \) defined above, is as follows
The dynamics of $\mathcal{P}$ defined above, implicitly define the dynamics of the belief states.

### 3.2.4 Evolutionarily Stable Policies

**Definition 3.19.** Let $D_{cn}(q_h, q_o, d, s)$ ($D$ for difference, $c$ for changed, $n$ for now) be the difference in expected total payoff resulting from being at a $q_h$ quality patch that didn’t change on the current round versus being at a formerly $q_o$ quality patch that did change this round. On the previous round foragers were either in belief state $(q_h, q_h, d^*(q_h, q_o), d^{t-1}, q_o, -1)$ or $(q_o, q_o, d^*(q_o, q_h), 1 - d^{t-1}, q_h, -1)$. The population
CHAPTER 3. DUCK PROBLEM 1

employs a policy of the standard form, so that 
\( d = d^*(q_h, q_o) \) and \( s = s^*(q_h, q_o) \).

\[
D_{cn}(q_h, q_o, d, s) = \left( \frac{q_h}{d} - \frac{1/\lambda}{1 - d} \right) + (1 - P_d) \left( (1 - \lambda)q_h(q_h, q_o, d, s) - (1 - \lambda)q_B(q_h, q_o, d, s) - 1 \right) (0)
\]
\[
+ (1 - P_d) \left( 1 - (1 - \lambda)q_h(q_h, q_o, d, s) \right) \left( \frac{s}{d} - \frac{q_h}{1 - s} \right)
\]
\[
+ (1 - P_d) \left( (1 - \lambda)q_h(q_h, q_o, d, s) - 1 \right) \left( 1 - \frac{s}{d} \right) \left( \frac{q_h}{d - s} - \frac{q^+}{1 - d + s} \right)
\]
\[
+ (1 - P_d)^2 I(q_{BB} > q_B) \left( (1 - \lambda)q_h(q_h, q_o, d, s) - 1 \right) \left( 1 - \frac{s}{d} \right) \left( \frac{q^+}{d - s} - \frac{q_h}{1 - d + s} \right)
\]

(79)

Where \( q^+ \) is the expected quality of a patch with quality greater than or equal to \( q_B \),
\( q^- \) is the expected quality of a patch with quality less than or equal to \( q_B \), and \( q^{++} \) is
the expected quality of a patch with quality greater than or equal to \( q_B \) but less than \( q_{BB} \).

Definition 3.20. Let \( D_{ho}(q_h, q_o, d, s) \) (\( D \) for difference, \( h \) for here, \( o \) for other) be
the difference in expected total payoff between a forager at a \( q_h \) quality patch, and a
forager at a \( q_o \) quality patch, given that neither patch changed quality on the previous
round, and that either patch might or might not change quality on the current round.

\[
D_{ho} = (1 - P_c) \left( \frac{q_h}{d} - \frac{q_o}{1 - d} \right) + (P_c/2)D_{cn}(q_h, q_o, d, s) - (P_c/2)D_{cn}(q_o, q_h, 1 - d, s)
\]

(80)

When \( D_{ho}(q_h, q_o, d, s) \) equals zero the expected total payoff of a forager at one
patch is the same as the expected total payoff of a forager at the other patch. Thus
if a population uses a policy such that for each quality pair \( (q_h, q_o) \),

\[
D_{ho}(q_h, q_o, d^*(q_h, q_o), s^*(q_h, q_o)) = 0,
\]
then there is no inherent advantage to being at one patch over another, provided that
neither patch has changed on the previous round.

**Definition 3.21.** Let $D_{cp}(q_h, q_o, d, s)$ ($D$ for difference, $c$ for changed, $p$ for previously) be the difference in expected total payoff resulting from staying at a $q_h$ quality patch that didn’t change this round or the previous round versus switching to a formerly $q_o$ quality patch that did change on the previous round. Thus on the previous round the foragers considered in this case were in belief state $(q_h, q_h, d^*, q_h, q_o, d, s, 1)$, (and the foragers at the other patch were in belief state $(q_h, q_o, 1 - d^*(q_h, q_o), q_o, d, t - 1, q_h, 2)$ with $q_o' \neq q_o$). The population employs a policy of the standard form, with $d = d^*(q_h, q_o)$ and $s = s^*(q_h, q_o)$.

$$D_{cp}(q_h, q_o, d, s) = \left( (1 - \lambda)q_h(q_h, q_o, d, s) - (1 - \lambda)q_h(q_h, q_o, d, s) - 1 \right) 0$$

$$+ (1 - (1 - \lambda)q_h(q_h, q_o, d, s)) \left( \frac{q_h}{1 - s} - \frac{q^+}{s} \right)$$

$$+ \left( (1 - \lambda)q_B(q_h, q_o, d, s)^{-1} \right) \left( \frac{q_h}{d - s} - \frac{q^+}{1 - d + s} \right)$$

$$+ (1 - P_d)I(q_{BB} > q_B) \left( (1 - \lambda)q_B(q_h, q_o, d, s)^{-1} - (1 - \lambda)q_B(q_h, q_o, d, s)^{-1} \right)$$

$$\left( \frac{q^+}{d - s} - \frac{q_h}{1 - d + s} \right)$$

(81)

**Definition 3.22.** Let $D_{ss}(q_h, q_o, d, s)$ ($D$ for difference, $s$ for stay and for switch) be the difference if expected total payoff of a forager that stays at the same patch as previously and a forager at that same patch that switches to the other patch.

$$D_{ss}(q_h, q_o, d, s) = (1 - P_c^s) \left( (1 - P_c) \left( \frac{q_h}{d} - \frac{q_o}{1 - d} \right) \right)$$

$$+ (P_c / 2) D_{cn}(q_h, q_o, d, s) - (P_c / 2) D_{cn}(q_o, q_h, 1 - d, s)$$

$$+ (P_c') D_{cp}(q_h, q_o, d, s)$$

(82)

When $D_{ss}(q_h, q_o, d, s)$ equals zero the expected total payoff of a forager that switches from a $q_h$ quality patch is the same as the expected total payoff of a forager
that stays at that $q_h$ quality patch. Thus if a population uses a policy such that for each quality pair $(q_h, q_o)$,

$$D_{ss}(q_h, q_o, d^*(q_h, q_o), s^*(q_h, q_o)) = 0,$$

then a forager can gain no advantage over its patch mates by switching or staying at a different frequency than its patch mates.

The criteria of an Evolutionarily Stable policy is that it can resist invasion by a rare mutant policy. Attention is restricted to policies of the standard form, so the form of a resident policy and the population distribution dynamics it will produce is well understood. In order to understand the fate of rare mutant policies, the form of these mutant policies must be similarly well understood. To this end, mutations are assumed to be small, and localized, so that a mutant policy will be identical to the resident policy, except that in one particular belief state, assumed to be of the form $b = (q_h, q_o, d^*(q_h, q_o), d^t, q_o, -1)$, the mutant will switch with a slightly higher or lower frequency than the resident, so that if $\pi_r$ is the policy of the resident and $\pi_m$ is the policy of the mutant, $\pi_m = \pi_r$ except for a single particular belief state, $b$, of the form described above where $\pi_m(b) = \pi_r(b) \pm \epsilon_s$, where $\epsilon_s > 0$ is some arbitrarily small positive constant.

**Definition 3.23.** Let a candidate policy, $\pi^*$ be any policy of the standard form with, target population distribution function $d^*$ and switching regime function $s^*$ such that for every combination of possible patch qualities $(q_h, q_o)$,

$$D_{ho}(q_h, q_o, d^*(q_h, q_o), s^*(q_h, q_o)) = 0$$

**Proposition 3.24.** All candidate policies are such that if they can be invaded by various mutant policies, then if a certain assumption, to be specified shortly, holds,
the subpopulations of mutant type foragers can only persist within the population at insignificant levels.

Proof. Since a mutant policy can only ever deviate from the resident policy in its response to a single belief state, without loss of generality we let the belief state where this deviation occurs be \( b = (q_h, q_h, d^*, d^{d-1}, q_o, -1) \). We let the resident policy, \( \pi^* \), be a candidate policy with \( s^* = s^*(q_h, q_o) \) and \( d^* = d^*(q_h, q_o) \), so that \( D_{ho}(q_h, q_o, d^*, s^*) = 0 \).

In the case where \( D_{ss}(q_h, q_o, d^*, s^*) < 0 \) and \( s^* < d^* \), then if the mutant forager has \( \pi_m(b) = \pi^*(b) - \epsilon_s \), and so switches from the \( q_h \) quality patch to the \( q_o \) quality patch more frequently than a resident type forager, then this mutant will have a slightly higher expected total payoff than resident type foragers, and so this mutant subpopulation can grow initially.

In the case where \( D_{ss}(q_h, q_o, d^*, s^*) < 0 \) and \( 0 < s^* \), then if the mutant forager has \( \pi_m(b) = \pi^*(b) + \epsilon_s \), and so switches from the \( q_h \) quality patch to the \( q_o \) quality patch less frequently than a resident type forager, then this mutant will have a slightly lower expected total payoff than resident type foragers, and so this mutant subpopulation cannot grow initially.

In the case where \( D_{ss}(q_h, q_o, d^*, s^*) > 0 \) and \( s^* > 0 \), then if the mutant forager has \( \pi_m(b) = \pi^*(b) + \epsilon_s \), and so switches from the \( q_h \) quality patch to the \( q_o \) quality patch less frequently than a resident type forager, then this mutant will have a slightly higher expected total payoff than resident type foragers, and so this mutant subpopulation can grow initially.

In the case where \( D_{ss}(q_h, q_o, d^*, s^*) > 0 \) and \( s^* < d^* \), then if the mutant forager has \( \pi_m(b) = \pi^*(b) - \epsilon_s \), and so switches from the \( q_h \) quality patch to the \( q_o \) quality
patch more frequently than a resident type forager, then this mutant will have a slightly lower expected total payoff than resident type foragers, and so this mutant subpopulation cannot to grow initially.

In the case $D_{ss}(q_h, q_o, d^*, s^*) = 0$ then the different frequency of switching that a mutant forager in belief state $b$ engages in compared to the resident type foragers at that same patch with that same belief state is selectively neutral. Thus it may be possible that a subpopulation of foragers using a rare mutant policy can grow initially.

In the cases where the mutant has a relative fitness advantages over its resident type patch mates, as a result of switching more or less frequently than the resident type, as the size of the subpopulation using the mutant policy increases this relative fitness advantage must decrease.

If the mutant foragers are switching more frequently from the $q_h$ quality patch than the resident type, and switching from the $q_o$ quality patch with the same frequency as the resident type, then as the size of the mutant subpopulation grows the proportions of the total populations at each patch, will be such that the proportion of the population at the $q_h$ quality patch, denoted $d_m$ will satisfy $d_m < d^*$. This means that $D_{ho}(q_h, q_o, d_m, s^*) < 0$, so that a forager at the $q_h$ quality patch will have a higher expected total payoff than a forager at the $q_o$ quality patch, given that the neither patch has changed quality on the previous round. The differential switching of the mutants and the residents also means that there will be, relative to the total proportion of mutant types, more mutant foragers at the $q_o$ quality patch, and fewer mutant foragers at the $q_h$ quality patch. This gives the mutant foragers a relative fitness disadvantage, and this disadvantage must increases as the size the mutant subpopulation increases. Note the implicit assumption here that although $d_m$ differs
from $d^*$ enough to affect the expected payoff of the foragers, this difference is not significant enough to impair a forager’s ability to infer the quality of the other patch from the movements of the population.

Similarly if the mutant foragers are switching less frequently from the $q_h$ quality patch than the resident type, this also gives the mutant foragers a relative fitness disadvantage which also must increases as the size of the mutant subpopulation increases.

It is assumed that the rate of decrease of the mutant’s relative advantage, and the rate of increase of the mutant’s relative disadvantage is such that before the proportion of the population using the mutant policy grows greater than some $\epsilon_m$, the expected total payoff of a forager using the mutant policy is less than the expected total payoff of a forager using the resident policy. Thus the mutant subpopulation can never grow beyond some proportion $\epsilon_m$ of the total population.

Note that this assumption is not needed in the case where $D_{ss}(q_h, q_o, d^*, s^*) = 0$, since the mutants start with no relative fitness advantage. This assumption is also not needed in the case where $D_{ss}(q_h, q_o, d^*, s^*) < 0$ and $s^* = d^*$ since it is not actually possible to do more switching than resident foragers are currently doing, nor in the case where $D_{ss}(q_h, q_o, d^*, s^*) > 0$ and $s^* = 0$ since it is not actually possible to do less switching than the resident foragers are currently doing.

Thus while these candidate policies are technically not evolutionarily stable strategies, or even evolutionarily convergent strategies, they are still of interest from an evolutionary perspective. Note that a candidate policy will be an ES policy if for each quality pair $(q_h, q_o)$, in addition to the requirements of being a candidate policy, one of the three following sets of conditions holds
1. \( D_{ss}(q_h, q_o, d^*, s^*) = 0 \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) = 0 \)

2. Without loss of generality \( d^* < 1 - d^* \), \( D_{ss}(q_h, q_o, d^*, s^*) < 0 \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) = 0 \) and \( s^* = d^* \)

3. \( d^* < 1 - d^* \), \( D_{ss}(q_h, q_o, d^*, s^*) > 0 \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) > 0 \) and \( s^* = 0 \)

In appendix B.2 there is a computer program which examines each possible patch quality pair, \((q_h, q_o)\), up to some given limit, and computes \( d^* \) and \( s^* \) for each patch quality pair so that if possible a \((d^*, s^*)\) pair satisfying, \( D_{ss}(q_h, q_o, d^*, s^*) = 0 \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) = 0 \) and \( D_{ho}(q_h, q_o, d^*, s^*) = 0 \) is found. If this is not possible a \((d^*, s^*)\) pair minimizing the sum of the absolute value \( D_{ss}(q_h, q_o, d^*, s^*) \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) \) is found, while ensuring \( D_{ho}(q_h, q_o, d^*, s^*) = 0 \). Interestingly in the these cases where \( D_{ss}(q_h, q_o, d^*, s^*) \neq 0 \) and \( D_{ss}(q_o, q_h, 1 - d^*, s^*) \neq 0 \), the minimizing pair \( d^* \) and \( s^* \) is always such that \( s^* = \min[d^*, 1 - d^*] \). The target population distribution for the policy computed in this way, \( d^*(q_h, q_o) \) is displayed in figure 2 as a function of the quality of the \( q_o \) quality patch for select values of \( q_h \), using parameters \( \lambda = 0.1, P_c = 0.1, P_d = 0.2 \).

### 3.3 Discussion

The fitness value of information is the increase in fitness from future decisions that are made differently as a result of that information, [7]. Thus a good way of assessing the value of information is to see what foragers using the ES or optimal strategy are willing to “pay” in terms of immediate fitness costs for that information. In the lone forager problem the value of knowing “the quality of the other patch” could be definitively calculated, and we find that the information’s value depends on how
likely the information is to be new, that is the degree to which it might change old beliefs or knowledge about the quality of the other patch. If the information is new its value further depends on how likely the information is to affect behaviour, and if the information does affect behaviour the value of the information further depends on the fitness difference resulting from the change in the forager’s behaviour as compared to the forager’s behaviour without the information. In the many foragers case, when a candidate policy was employed switching between patches was free. This made the problem more tractable, but also meant that information was free for the foragers, which prevents us from assessing what the value of knowing “the quality of the other patch” was for the foragers in the many forager problem. This makes a comparison between the value of information in the two cases unfruitful, and so we modify the problem.
Chapter 4

Duck Problem 2

To remedy the shortcoming of the previous pair of problems, in the following pair of problems I introduce an explicit cost to switching from one patch to the other. I also give foragers the option of observing the other patch from afar. Inspired by my own experiences watching ducks move from one group of people dispensing bread at one area of a pond, to another group dispensing bread at another area, I envision the costs of switching and observing in terms of forgone opportunity costs and hence the costs of switching and observing are proportional to the payoff at the relevant patch. See figure 3 for further details. I set the costs of observing as substantially less than the costs of switching, the idea being that looking over at a distant patch is less costly than actually moving to that distant patch. The costs of observation and switching could just as easily be fixed instead of varying with patch quality with no great effect on the model. Similarly the option of observation is not strictly necessary, nor does observation necessarily need to be less costly than switching. With no observation foragers would be forced to switch and sample the other patch to gain information. The advantage of giving information gathering and migration distinct actions is that
there is less ambiguity about why a forager is choosing a particular action. To simplify
the problem further I initially investigate this problem with a fixed terminal time,
and with no environmental change. While this makes the problem relatively simple
it provides a basic framework for developing the techniques that will be helpful when
dealing with the more complicated unknown duration, dynamic environment problem.

4.1 Lone Forager, Costly Switching,
   Costly Observation, Fixed Terminal Time,
   Static Environment

In this problem we suppose that there is a single forager who plays the following game
for \( T \) rounds. The forager begins the game on patch \( x \) of patches \( x \) and \( y \). At the
beginning of the game payoffs for each patch, \( q_x \) and \( q_y \), are drawn independently from
the same geometric distribution with parameter \( \lambda \). Once drawn these payoffs remain
fixed throughout the game. The forager is initially aware of the quality of its starting
patch, \( q_x \), and unaware of the quality of the other patch, \( q_y \). On each round the
forager must choose between one of three options: exploiting the patch it is currently
located at, observing the value of the other patch, or switching to the other patch.
If the forager exploits a patch it receives the full payoff associated with that patch
for that round. If the forager observes at a patch it receives a proportion \( c_o \) (\( c \) for
cost although this is not actually the cost, \( o \) for observation) of the payoff associated
with that patch for that round. If the forager switches from one patch to the other
it receives a proportion \( c_s \) (\( c \) for cost, \( s \) for switching) of the payoff associated with
the patch it has just switched to.
Figure 3: $c_o$ and $c_s$ are not the costs of observing and switching but rather the factor by which payoff is reduced when these actions are employed. Thus the cost or loss due to observation from patch $x$ is $q_x(1 - c_o)$ and the loss or cost due to switching to patch $y$ is $q_y(1 - c_s)$, relative to already being at the $y$ patch, and $q_x - (q_y \cdot c_s)$ relative to staying at the $x$ patch.

A quick inspection of the game makes it clear that there are only three possibilities for an optimal strategy:

1. Exploiting for all $T$ rounds at patch $x$ and ignoring patch $y$.

2. Observing the value of patch $y$ on the first round, and contingent on the results of that observation switching to patch $y$ or staying at patch $x$, and exploiting ever after.

3. Switching to patch $y$ immediately and then contingent on the quality of patch $y$, exploiting patch $y$ ever after or switching back to patch $x$ and exploiting patch $x$ ever after.

Though this problem can be framed as a Bayesian sequential decision problem, and in fact the techniques of Bayesian sequential decision theory are implicitly used
its solution, the use of the formal machinery of that theory would serve primarily
to obscure the relatively simple calculations and reasoning required to compute an
optimal decision rule for a forager in the situation described above.

**Problem 4.1 Summary**

- The environment is static.
- The quality of the unknown patch is drawn from a geometric distribution
  with parameter \( \lambda \).
- The foraging session lasts for \( T \) rounds.
- Moving from one patch to another is costly, and this cost is proportional to
  the payoff at the patch being moved to.
- Observing one patch from the other is costly, though less so than moving
  between patches. This cost is in proportion to the payoff at the patch from
  which the observation is made.
- A forager knows
  - The quality of its starting patch.
  - The parameters of the model, \( T, \lambda, c_o \) and \( c_s \).
- A forager does not know
  - The quality of the other patch unless it has observed or switched.

**What we are looking for:** Whether the forager should exploit, observe or switch
on the first round.

The three possibilities above correspond to the following cases. For a sufficiently
high starting patch quality, \( q_x \), the probability of the other patch’s payoff, \( q_y \), being
better is small so observation is a waste. For intermediate \( q_x \), it is wise to bet hedge
and observe \( q_y \) before committing to a patch. For sufficiently small values of \( q_x \) the
probability of \( q_y \) being sufficiently greater than \( q_x \) is large enough that it is again
a waste to observe, and the forager should just switch. The dilemma faced by a
forager is that of determining what constitutes a sufficiently high, intermediate and
sufficiently low initial patch quality. To do this the forager must compare the expected
payoffs resulting from choosing to either exploit, observe, or switch on the first round, and then proceeding in an optimal fashion.

The expected payoff when exploiting initially is

\[ q_x \cdot T \]  

The expected payoff when observing initially is

\[ q_x c_o + \left( \left\lfloor \frac{q_x(T - 1)}{T - 2 + c_s} \right\rfloor + \frac{1}{\lambda} \right) (T - 2 + c_s) P_{\text{switch}} + q_x(T - 1)(1 - P_{\text{switch}}) \]  

Where

\[ P_{\text{switch}} = (1 - \lambda) \left\lfloor \frac{q_x(T - 1)}{T - 2 + c_s} \right\rfloor \]  

is the probability that the \( y \) patch will be worth switching to and \( 1 - P_{\text{switch}} \) is the probability that it is better to stay at the \( x \) patch, given the costs of switching and the time left to recoup those costs.

The expected payoff when switching initially is

\[ \frac{c_s}{\lambda} + \left( \left\lfloor \frac{q_x(T - 2 + c_s)}{T - 1} \right\rfloor + \frac{1}{\lambda} \right) (T - 1) P_{\text{stay}} + q_x(T - 2 + c_s)(1 - P_{\text{stay}}) \]  

Where

\[ P_{\text{stay}} = (1 - \lambda) \left\lfloor \frac{q_x(T - 2 + c_s)}{T - 1} \right\rfloor \]  

is the probability that after switching to the \( y \) patch it is worth staying there and \( 1 - P_{\text{stay}} \) is the probability that after to switching to the \( y \) patch, it is of a low enough value that it is worthwhile to switch back to the \( x \) patch, given the costs of switching and the time left to recoup those costs.

Examining these expressions gives a sense of how the various parameters affect the value of observation. As before the value of knowing “the quality of the other patch” depends on how likely that information is to change behaviour, and what
the relative gains might be from that change in behaviour. When the value of the forager’s starting patch is very low or very high, then the chance that the information it receives about the \( y \) patch will affect its choice of where to exploit is low, making switching or exploiting, as appropriate, preferable to observation. As the duration of the foraging session, \( T \) increases, the potential differences between an informed versus an uninformed choice becomes more drastic, making ignoring the \( y \) patch a riskier proposition.

Using expressions 84, 85 and 87, it is easy to calculate the optimal strategy for any specific set of parameters. A program that does this can be found in appendix B.3. The optimal strategy is conveniently described by a double threshold which gives the minimum starting patch quality for which observation dominates switching and exploiting, and the minimum starting patch quality for which exploiting dominates observation. Table 1 gives these threshold values for a variety of values of \( T \) and \( \lambda \) when \( c_o = 0.5 \) and \( c_s = 0.9 \). For instance suppose a forager knows that the quality of the other patch is distributed geometrically with parameter \( \lambda = 0.1 \) and that the foraging session will last for 50 rounds. If the quality of the starting patch is 5 or less then it is likely a waste of time to observe and the forager should just switch to the other patch. If the quality of the starting patch is 6 or greater, but less than 45, then it is best to play conservatively and observe before choosing where to exploit. If the quality of the initial patch is 45 or greater then the chance of the other patch being

<table>
<thead>
<tr>
<th>( T )</th>
<th>( \lambda = 0.05 )</th>
<th>( \lambda = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11.62</td>
<td>6.30</td>
</tr>
<tr>
<td>25</td>
<td>11.79</td>
<td>6.39</td>
</tr>
<tr>
<td>50</td>
<td>11.91</td>
<td>6.45</td>
</tr>
<tr>
<td>100</td>
<td>11.104</td>
<td>6.50</td>
</tr>
</tbody>
</table>
better is so slight, and the cost of observation so great that observation is a waste of time and the forager should just exploit.

4.2 Many Foragers, Costly Switching, Costly Observation, Fixed Terminal Time, Static Environment

We now turn to the many forager analogue of the previous problem. We assume that the entire population is initially at one of the patches, but not aware of the quality of the other patch. The base payoff per round to a forager is the quality of its current patch divided by the proportion of the population currently at that patch. This base payoff is then scaled by the costs of observation and switching as appropriate. We are looking for an ES strategy which prescribes the probability with which a forager should exploit, observe, or switch, on each round, given which patch they are at, how many other foragers are at that patch with them, and how many rounds have passed.

In this problem a forager that does not know the quality of the other patch must balance a tangle of frequency dependent risks and costs, by carefully choosing between exploiting, observing, or switching blindly. A forager exploits in the hope that there will be enough other foragers migrating and observing that the population distribution will be brought into equilibrium at no cost to the exploiters. A forager switches in the hope that the other patch will be of high enough quality, that the costs of switching will be offset by the high quality of that patch. A forager observes in the hopes that the other patch will be of high enough quality that even after all of the observers and all of the switchers have gone over to the other patch the quality of the other
patch is great enough to offset the cost of switching, and compensate for the cost of observation. The rewards for engaging in each action are then very much dependent not only on how many foragers will engage in each action in the present, but also on many foragers will engage in the various actions on subsequent rounds.

**Problem 4.2 Summary**

**Assumptions:**

- The environment is static.
- The quality of the unknown patch is drawn from a geometric distribution with parameter $\lambda$.
- The foraging session lasts for $T$ rounds.
- Moving from one patch to another is costly, and this cost is proportional to the payoff at the patch being moved to.
- Observing one patch from the other is costly, though less so than moving between patches. This cost is in proportion to the payoff at the patch from which the observation is made.
- Foragers can make inferences about the quality of the other patch based on the switching of their patch mates.
- The population is large enough that the proportion of the population employing a given action on a round is effectively deterministic, even if the action employed by any one individual is the stochastic result of a mixed strategy.
- A forager knows
  - The quality of its starting patch.
  - The parameters of the model, $T, \lambda, c_o$ and $c_s$.
  - The proportion of the population at each patch.
  - The policy employed by the population from which inferences can be made about the quality of the other patch
- A forager does not know
  - The quality of the other patch until it has observed or switched, or equilibrium is achieved.

**What we are looking for:** Whether a forager should exploit, observe or switch on each round.
4.2.1 POMDP Formulation

As in the previous chapter the standard POMDP formulation can be augmented with the tools of evolutionary game theory to precisely describe this problem. $h$ (for here) and $o$ (for other) are used to denote the patch a forager currently occupies and the other patch respectively in the case where a subjective description of the state is appropriate, and $x$ and $y$ are used to denote the patch that the population initially starts at and the other patch respectively in the case where an objective description of the state is appropriate. Elements of the environmental state space $S$ consist of the current quality of the two patches, $q_h$ and $q_o$ (or $q_x$ and $q_y$ if appropriate) and the proportion of the population at each patch $d_h$ and $d_o = 1 - d_h$ (or $d_x$ and $d_y$ if appropriate). Thus $s \in S$ is of the form $s = (q_h, q_o, d_h)$ or $s = (q_x, q_y, d_x)$. The action space is $A = \{ -1, 0, 1 \}$, where $a = 1$ indicates a forager exploiting the same patch as on the previous round and $a = 0$ indicates a forager observing the other patch, and $a = -1$ indicates a forager switching from its current patch to the other patch.

The patch qualities are static so the only change in the objective state comes through the movement of the foragers between the patches.

$$
\begin{align*}
    d_x^{t+1} &= \Pr(a = 1 \text{ or } a = 0 | \text{forager at } x \text{ patch})(d_x^t) \\
    &\quad + \Pr(a = -1 | \text{forager at } y \text{ patch})(1 - d_x^t) \\
\end{align*}
$$

Note the implicit assumption here that the population is large enough that the dynamics of the population distribution can be treated as deterministic with respect to expectations of the proportion of foragers at a given patch that will take a given action, despite the fact that an individual forager’s action may be stochastically chosen.
The changes in the subjective state are only slightly more complicated.

\[
d_h^{t+1} = \Pr(a = 1 \text{ or } a = 0 | \text{forager at } h \text{ patch}) (d_h^t)
\]

\[+ \Pr(a = -1 | \text{forager at } o \text{ patch}) (1 - d_h^t) \]

if \(a = 1 \text{ or } a = 0\) \hspace{1cm} (90)

\[
d_h^{t+1} = \Pr(a = -1 | \text{forager at } h \text{ patch}) (d_h^t)
\]

\[+ \Pr(a = 1 \text{ or } a = 0 | \text{forager at } o \text{ patch}) (1 - d_h^t) \]

if \(a = -1\) \hspace{1cm} (91)

The reward function for a forager employing action \(a \in \{-1, 0, 1\}\) after experiencing state \(s^t = (q_h, q_o, d_h^t)\) and then transitioning to state \(s^{t+1} = (q_h', q_o', d_h^{t+1}, I_p')\) is

\[
R_{-1}(s^t, s^{t+1}) = c_s \cdot q_h' / d_h^{t+1}
\]

\[
R_0(s^t, s^{t+1}) = c_o \cdot q_h' / d_h^{t+1}
\]

\[
R_1(s^t, s^{t+1}) = q_h' / d_h^{t+1}
\]

Recall that \(c_o\) is the scaling factor by which the payoff of an observer is reduced and \(c_s\) is the scaling factor by which the payoff of a migrating forager is reduced.

The objective of an individual forager is to optimize the following

\[
\sum_{t=1}^{T} R_a(s^t, s^{t+1})
\]

where \(T\) is an integer known to the foragers. A forager does this by adopting a policy function, \(\pi : \mathcal{B} \rightarrow [0, 1]\), which gives the probabilities with which a forager should exploit, observe or switch given the current belief state \(b \in \mathcal{B}\) of the forager. Since the forager does not always know the state of the system precisely, \(b\) is a representation
of all of the information that a forager can use to inform its decision of which action to take.

Every round the forager receives information in the form of $i \in \mathcal{O}$. When a forager takes action $a$ after experiencing state $s^t = (q_h, q_o, d_h^t)$ and transitions to state $s^{t+1} = (q_h', q_o', d_h^{t+1})$ the forager receives information $i^{t+1} = (q_h, d_h^{t+1}, t = 1)$, that is the quality of the patch it is currently at, the current distribution of the population, and the number of rounds that have passed in the foraging session. The forager’s policy must be based on information that can be inferred from the history of this data. Thus all the foragers in the population always know: the quality of the starting patch $q_x$, the current distribution of the population $d_h$, how many rounds have passed in the foraging session, $t$, and thus how many rounds are left in the session $T - t$, and their current location. It is assumed that all of the foragers also know the policy employed by the population. The only important piece of information that not all of the foragers know (at least initially) is the quality of the patch that they do not start at, $q_y$. If a forager has at any time in the session switched or observed then it will know $q_y$ with certainty. Depending on the policy employed by the population, a forager may or may not be able to infer information about the quality of $q_y$ without ever having observed or switched. As in the previous chapter the dynamics and the form of the belief states are then dependent on the policy adopted by the population. Thus we must restrict our attention to policies of a certain form before precisely characterizing the belief states and their dynamics.
4.2.2 The Form of ES Policies

In this section we seek to characterize the form of an ES policy. To this end I will introduce conditions and prove that they are necessary for a policy to be ES. Once we have proved that a condition is necessary, it will be assumed in all of the subsequent claims and their proofs that only policies for which the previously established conditions hold are being considered.

**Definition 4.1.** We say there is incentive to switch from a patch of quality $q_h$ with $d^t_h$ of the foragers anticipated to be at that patch to a patch of quality $q_o$ with $d^t_o = 1 - d^t_h$ of the foragers anticipated to be at that patch, on round $t$, when there are $T + 1 - t$ rounds remaining in the session, if

$$(T + 1 - t)(q_h/d^t_h) < (T + c_s - t)(q_o/d^t_o)$$

(95)

This means that if the proportion of the foragers at the $o$ patch does not increase enough to reverse this inequality in future rounds, then a forager switching to the $o$ patch will have a higher total payoff than a forager staying at the $h$ patch despite having to pay the cost of switching from one patch to another.

**Definition 4.2.**

$$g(t) = 1 - (1 - c_s)/(T + 1 - t)$$

(96)

denotes the minimum factor by which the payoff at the patch being switched to must exceed the payoff at the patch being switched from on round $t$ of the foraging session.

In other words, for there to be incentive to switch to a patch with payoff $q_o/d_o$ from a patch with payoff $q_h/d_h$ on round $t$ it must be the case that

$$g(t) \cdot q_o/d_o > q_h/d_h$$

(97)
or equivalently

\[ d_h > \frac{q_h}{q_h + q_o \cdot g(t)} \]  \hspace{1cm} (98)

**Condition 4.3.** When a forager knows enough about \( q_o \) and \( d^h_t \) that the forager knows whether or not the following inequality holds

\[ d^h_t \geq \frac{q_h}{q_h + q_o \cdot g(t)} \]  \hspace{1cm} (99)

and the forager knows that after round \( t \) the distribution of the population won’t change enough to change the evaluation of inequality 99, then the forager will switch from the \( h \) patch to the \( o \) patch, on round \( t \) only if it knows that this inequality does hold. Further if the forager knows that this inequality holds strictly then the forager will certainly switch.

**Proof.** Consider a policy that does not satisfy this condition, \( \pi_{\bar{c}} \). Thus there is a situation where a forager using \( \pi_{\bar{c}} \) knows that

\[ d^h_t < \frac{q_h}{q_h + q_o \cdot g(t)} \]  \hspace{1cm} (100)

and switches to the \( o \) patch.

Consider a policy \( \pi_c \) which is identical to \( \pi_{\bar{c}} \) in every way except that in the cases described above where a forager using \( \pi_{\bar{c}} \) will switch, and hence violate the condition a forager using \( \pi_c \) will not switch, and hence satisfy the condition. Clearly then foragers using \( \pi_{\bar{c}} \) will have an expected total payoff which is strictly less than \( \pi_c \).

An identical argument can be applied to the case where a forager using \( \pi_{\bar{c}} \) knows

\[ d^h_t > \frac{q_h}{q_h + q_o \cdot g(t)} \]  \hspace{1cm} (101)

and stays at the \( h \) patch.

This condition ensures that informed foragers switch to the other patch only if
the cost of switching will be compensated for by the higher per round payoff at the other patch over the course of the time remaining in the foraging session, and that if the cost of switching will be more than compensated for by the higher per round payoff at the patch being switched to an informed forager will certainly switch there.

**Condition 4.4.** When a forager knows enough about $q_o$ and $d_h$ that the forager knows whether or not inequality 99 holds, and the forager knows that the population distribution will not change enough to alter the evaluation of inequality 99 then a forager will only ever exploit or switch, but never observe.

*Proof.* If a forager already knows enough about the patch qualities and the population distribution to know whether or not switching will result in a higher payoff than staying, if a forager observes in this case it will be in precisely the same situation as if it had exploited, except that it will have paid the cost observation, thus observation is strictly dominated by exploiting in this situation and a forager trying to maximize its total payoff need never observe in this situation. □

This condition ensures that foragers never make redundant observations.

**Definition 4.5.** The population is said to be in equilibrium, if no forager knowing enough about $q_o$ and $d_h$ to evaluate inequality 99 will switch from one patch to the other, and if all foragers know enough about $q_o$ and $d_h$ to evaluate these inequalities.

Note that if the population is in equilibrium the distribution of the population will not change. Also note that in the previous chapter where there was no explicit cost to switching there was a unique equilibrium population distribution for a given pair of patch qualities, whereas in this problem where there are explicit costs to switching
between patches there is a continuous range of population distributions that satisfy the equilibrium condition.

**Definition 4.6.** For a policy $\pi$ employed by the population let $S^t$ and $O^t$, called switching and observation quotas, denote the proportion of the total population, at the $x$ patch, that will switch or observe respectively on the $t^{th}$ round of the foraging session given that those foragers at the $x$ patch do not know whether inequality 99 holds or not.

**Definition 4.7.**

\[
P^1_s = \min \left[ 1, \max \left[ 0, \left( \frac{1}{O^{t-1}} \right) \left( d^{t-1}_x - \frac{q_x}{q_x + q_y \cdot g(t)} - S^t \right) \right] \right]
\]

(102)

is the proportion of the foragers that have observed the $y$ patch from the $x$ patch on round $t - 1$ that should switch to the $y$ patch so that the total future payoff of the foragers that observe on round $t - 1$ and then switch to the $y$ patch on round $t$ (if there are any), is the same as the total future payoff of the foragers that observe on round $t - 1$ and stay and exploit at the $x$ patch on round $t$ (if there are any).

\[
P^2_s = \min \left[ 1, \max \left[ 0, \left( \frac{1}{d^{t-1}_y} \right) \left( d^{t-1}_y + S^t - \frac{q_y}{q_y + q_x \cdot g(t)} \right) \right] \right]
\]

(103)

is the proportion of the foragers that are at the $y$ patch on round $t - 1$ that should switch to the $x$ patch on round $t$ so that the total future payoff of the foragers at the $y$ patch on round $t - 1$ that then switch back to the $x$ patch on round $t$ (if there are any), is the same as the total future payoff of the foragers at the $y$ patch on round $t - 1$ and that stay at the $y$ patch on round $t$ (if there are any). If there are foragers at the $x$ patch that are still unable to evaluate inequality 99 with certainty on round $t - 1$. 

**Condition 4.8.** If a forager has observed on the previous round, \( t - 1 \), it will switch to the \( y \) patch on round \( t \) with probability \( P_1^s \).

If a forager is at the \( y \) patch on round, \( t - 1 \), it will switch back to the \( x \) patch on round \( t \) with probability \( P_2^s \), if there are foragers at the \( x \) patch that are still unable to evaluate inequality 99 with certainty on round \( t - 1 \).

**Proof.** Since the foragers are assumed to know the policy employed by the population, including \( S^t \), \( O^t \), and the probability with which foragers that have observed the other patch and found it to be of a certain value will switch to that other patch, then a forager that knows the quality of the other patch can deduce the population distribution in the next round. Thus the only way for condition 4.3 to be satisfied is for the switching probabilities of observers to be as described in condition 4.8. □

This condition ensures that the proportion of informed foragers that switches from one patch to another is such that total future payoffs of foragers that stay and exploit (if there are any) and the total future payoff of the foragers that switch patches (if there are any) is the same. For ease of reference we state this as a corollary.

**Corollary 4.9.** Any two foragers at the same patch that are informed about the quality of the other patch have the same total future payoff.

**Corollary 4.10.** The population is not in equilibrium if and only if all of the foragers at the \( x \) patch that have observed the quality of the \( y \) patch on round \( t - 1 \) or earlier have switched to the \( y \) patch by round \( t \).

**Proof.** Since condition 4.8 holds, if fewer than all of the observers have switched to the \( y \) patch on round \( t \), the non-observing foragers at the \( x \) patch can deduce that
the quality of the $y$ patch is such that

$$d^t_x \leq \frac{q_h}{q_h + q_o \cdot g(t)} \quad (104)$$

and so any subsequent switching or observing would be pointless, since

$$\frac{q_h}{q_h + q_o \cdot g(t)} < \frac{q_h}{q_h + q_o \cdot g(t + 1)} \quad (105)$$

\[ \square \]

**Corollary 4.11.** Given a policy with $S^i$ and $O^i$, if the population is not in equilibrium by round $t$ the distribution of the population is

$$d^t_x = 1 + O^t - \sum_{i=1}^{t} S^i + O^i \quad (106)$$

We denote this non equilibrium proportion of the population at the $x$ patch as $d^*(t)$

**Proof.** This is the proportion of the population at the $x$ patch on round $t$ given that all of the foragers that have made observations prior to round $t$ have had incentive to switch to the $y$ patch, which by corollary 4.10 is only the case if the population is not in equilibrium.

\[ \square \]

**Condition 4.12.** The probability of a forager switching to patch $y$, on the second round, after exploiting on the first round is zero.

**Proof.** Given the value of $q_y$ and the policy employed by the population, one of the following two situations must be the case.

1. There is incentive to switch to the $y$ patch on the second round, that is

$$\frac{T - 1}{1 - S^1} \frac{q_x}{q_y} < \frac{T - 2 + c_s}{S^1} \frac{q_y}{S^1}$$

In this case the payoff a first round switche receives on the first round is greater than the payoff a second round switche receives on the second round

$$c_s \cdot q_y / S^1 > c_s \cdot q_y / d^2_y$$
and the payoff a first round switcher receives on the second round is greater than the payoff a second round switcher receives on the first round

\[ q_y/d_y^2 > q_x/d_x^2 > q_x/(1 - S^1). \]

Thus a forager switching on the first round has higher total payoff than a forager switching on the second round.

2. There is not incentive to switch to the \( y \) patch on the second round, that is

\[ (T - 1) \frac{q_x}{1 - S^1} > (T - 2 + c_s) \frac{q_y}{S^1}. \]  \hspace{1cm} (108)

Then a forager switching on the second round will have a lower total payoff than a forager that exploits on the second round.

Consider a policy that has some positive probability of switching to the \( y \) patch on the second round after exploiting at the \( x \) patch on the first round. Given that one of the two cases above must occur, this policy can be improved, either by decreasing the probability of switching after exploiting to zero and increasing the probability of exploiting after exploiting, or by increasing the probability of switching on the first round while decreasing the probability of exploiting on the first round, or some combination of these two options.

Lemma 4.13. A forager that exploits at the \( x \) patch until round \( t - 1 \) and then observes on round \( t \) will only have a higher total payoff than a forager that exploits at the \( x \) patch until round \( t \) in the case where \( q_y \) is high enough that

\[ (T + 1 - t) \frac{q_x}{d^*(t)} < (T - t + c_s) \frac{q_y}{1 - d^*(t)}. \]  \hspace{1cm} (109)

Proof. If it is not the case that

\[ (T + 1 - t) \frac{q_x}{d^*(t)} > (T - t + c_s) \frac{q_y}{1 - d^*(t)}. \]  \hspace{1cm} (110)
then the population will be in an equilibrium distribution either on the round $t$ when
the observations take place, or on round $t+1$, by corollary 4.10 and condition 4.8.
Further this means that at least some of the foragers that observed on round $t$ stay
at the $x$ patch and exploit and by corollary 4.9, of those foragers that observed on
round $t$ those that switch and those that stay have the same total payoff. Now the
foragers that observe on round $t$ and stay at the $x$ patch and exploit ever after have
the same total payoff as those foragers that exploited on round $t$, and will continue
to exploit there ever after, less the cost of making an observation on round $t$.  

Note that, by corollary 4.10 and condition 4.3, there will only ever be a forager
observing on round $t$ if all of the foragers that have observed on previous rounds have
switched to the other patch, and so the population is not in equilibrium as of round
$t-1$.

**Condition 4.14.** The probability of a forager observing on the second round, after
exploiting on the first round is zero.

*Proof.* Given the value of $q_y$ and the policy employed by the population, one the
following two situations must be the case.

1. There is incentive for all of the foragers that observed on the second round to
switch to the $y$ patch on the third round, that is

\[ (T - 1) \frac{q_x}{d^*(3)} < (T - 2 + c_s) \frac{q_y}{1 - d^*(3)} \]  

(111)

Note that in this case there will certainly have been incentive for all of the
foragers that observed on the first round to switch to the $y$ patch on the second
round. In this case the payoff a first round observer receives on the second
round is greater than the payoff a second round observer receives on the third round

\[ c_s \cdot q_y/(1 - d^*(2)) > c_s \cdot q_y/(1 - d^*(3)) \]

and since

\[ q_x/(d^*(1)) < q_x/(d^*(2)) < q_y/(1 - d^*(3)) \]

the payoff a first round observer receives the first and third rounds combined is greater than the payoff a second round observer receives on the first and second rounds combined that is

\[ (q_x/d^*(1))(c_o - 1) + q_y/(1 - d^*(3)) - c_o \cdot q_x/(d^*(2)) > 0 \]

Thus in this case a forager observing on the first round has higher total payoff than a forager observing on the second round.

2. There is not incentive for all of the foragers that observed on the second round to switch to the \( y \) patch on the third round, that is

\[ (T - 1)\frac{q_x}{d^*(3)} > (T - 2 + c_s)\frac{q_y}{1 - d^*(3)} \]  \hspace{1cm} (112)

Then by lemma 4.13 the foragers that observe on the second round will always have a lower total payoff than the foragers that exploit on the second round in this case.

Consider a policy that has some positive probability of observing on the second round. Given that one of the two cases above must occur, this policy can be improved, either by decreasing the probability of observing after exploiting to zero and increasing the probability of exploiting after exploiting, or by increasing the probability of observing on the first round while decreasing the probability of exploiting on the first
round, or some combination of these two options.

There is one final condition that we would like to place on the form of candidate ES policies, however the proof of this final condition will be much clearer after some discussion of the population distribution dynamics.

4.2.3 Population Distribution Dynamics

Assuming that the population is using a policy that satisfies all of the above conditions, the population can come into an equilibrium distributions in one of three ways.

1. An informed group of foragers can switch from the $y$ patch to the $x$ patch. In this case at equilibrium

\[
d_y = \frac{q_y}{q_y + q_x \cdot g(t)}
\]

and thus the payoff per forager per round at the $y$ patch is $q_y/d_y = q_y + q_x \cdot g(t)$ which is slightly lower than the payoff per round per forager at the $x$ patch which is $q_x/d_x = q_y/g(t) + q_x$. This type of equilibrium comes about when too many foragers have switched to the $y$ patch. Typically the foragers already at the $y$ patch will anticipate that too many foragers will be joining them and preemptively switch patches on the very turn that an excess of migrant foragers arrive at the $y$ patch from the $x$ patch. The exception to this is when there have been too many switchers on the very first round. In this situation there are no foragers yet at the $y$ patch to compensate for the excess of $x$ to $y$ migrations. Thus if there is an excess of $x$ to $y$ migrations on the first round, equilibrium is achieved the round after, where as if there is an excess of $x$ to
y migrations on subsequent rounds, equilibrium is achieved on that very round by the preemptive back migration of the foragers already at the y patch.

2. An uninformed group of foragers can switch from the x patch to the y patch and upon arrival find that neither they at the y patch have incentive to switch back, nor do the foragers at the x patch have incentive to switch to the y patch. In this case at equilibrium we have that

$$d_x < \frac{q_x}{q_x + q_y \cdot g(t)}$$

(114)

and

$$d_y < \frac{q_y}{q_y + q_x \cdot g(t)}$$

(115)

3. An informed group of foragers can switch from the x patch to the y patch. In this case at equilibrium

$$d_x = \frac{q_x}{q_x + q_y \cdot g(t)}$$

(116)

and thus the payoff per forager per round at the x patch is $q_x/d_x = q_x + q_y \cdot g(t)$ which is slightly lower than the payoff per round per forager at the y patch which is $q_y/d_y = q_x/g(t) + q_y$. This sort of equilibrium comes about when not enough foragers have switched to the y patch on previous rounds but enough foragers at the x patch have observed the value of the y patch on the preceding round so that only some of those observers can switch to the y patch before there is no longer any incentive to switch to the y patch.

With this understanding we are now able to prove the final condition on ES policies.
Condition 4.15. A forager never observes on even numbered rounds, and only switches on even rounds if it has observed on the previous round.

Proof. If equilibrium is not achieved on the second round, since all of the foragers that observed on the first round, switched to the $y$ patch on the second round, then on the third round the remaining foragers at the $x$ patch, none of which has observed the $y$ patch, are faced with a decision problem nearly identical to that faced by all of the foragers on the first round of the foraging session. As on the first round these foragers know that the quality of the other patch is great enough that some of the foragers at the $x$ patch potentially have incentive to switch there, but these foragers do not know the precise quality, $q_y$, of the other patch. The arguments used to prove condition 4.12 and condition 4.14, can then be applied to show that these foragers that remain at the $x$ patch on round 2, should never observe on round 4 and should only switch on round 4 if they have first observed on round 3 and found the $y$ patch to be worth switching to. Condition 4.15 follows inductively.

Definition 4.16. A policy which satisfies all of conditions 4.3, 4.4, 4.8, 4.12, 4.14, and 4.15, is called a candidate policy.

Thus when a candidate policy is employed the game proceeds in two round steps. On the first of these two rounds, some of the foragers at the $x$ patch switch, some observe and some exploit. On the second round, if equilibrium hasn’t already been achieved by the switchers, then if possible the informed foragers who observed on the first round bring the population distribution to equilibrium. If there were too few observers and switchers to bring the distribution to equilibrium then the game proceeds to the next two round step, where the foragers remaining at the $x$ patch
face a decision problem which is identical in form to the decision problem faced on the first round, though the specifics of the problem have changed.

Which equilibrium type occurs and when depends on the strategy employed by the population and the patch qualities $q_x$ and $q_y$. Recall from corollary 4.11 that if the population is not at equilibrium on round $t$ then necessarily

$$d_x^t = d_x^*(t)$$

and $1 - d_x^t = d_y^t$. Figure 4 shows how the equilibrium type is determined by a population’s policy and the value of $q_y$.

![Figure 4: The equilibrium type achieved and the round that it is achieved on as a function of the true value of $q_y$ and the strategy of the population as manifest in $d_x(t)$ and $d_y(t)$.

The ES foraging strategy specifies with what probability a forager should exploit, observe, or switch on each round, so that the expected value for a forager using any of these actions is the same. However the value of exploiting, observing, and switching on each round is not only contingent on the actions of the population on that round, but also contingent of the actions of the population on subsequent rounds.

Now that the form of the policies employed by the foragers is better understood the belief states of the foragers can be characterized. The belief state of a forager has the form $b = (q_h, \hat{q}_o, d_h^t, t, a^t)$. Where $q_h \in \mathbb{Z}^+$ is the quality of the forager’s current patch, $\hat{q}_o \in \mathbb{Z}$ conveys the information that a forager has about the quality of the patch that it is not currently at. If $\hat{q}_o < 0$, then $\hat{q}_o$ is the negative of the lower bound (inclusive)
a forager can place on the quality of the other patch. If $\hat{q}_o > 0$ the forager knows the that the quality of the other patch is $\hat{q}_o$ with certainty. If $\hat{q}_o = 0$ then although the forager does not know the precise value of the other patch it can infer that the other patch is not worth switching to (and so also not worth observing). $d_h \in [0, 1]$ is the proportion of the total population at the foragers current patch. $t \in \mathbb{Z}^+$ is the number of rounds that have passed in the foraging session. $a^t \in \{-1, 0, 1\}$ is the action that forager performed on round $t$ based on the information that it had after round $t - 1$ had passed. With this belief state characterization, and knowing that the policy employed by the population is a candidate policy, it is not difficult to compute the transition probabilities and costs from one state to another for a given action, and these can be found in appendix A.3.

Given a population employing a candidate policy, $\pi$, the dynamics of the population distribution are fixed by the (initially) unknown value of $q_y$.

**Definition 4.17.** Let $V_\pi(b)$ be the expected total future payoff of a forager in belief state $b$ using policy $\pi$, in a population using policy $\pi$.

Let $V_{\pi^{-1}}^b(b)$ be the expected total future payoff of a forager in belief state $b$ that takes the action of switching patches, when in belief state $b$ and then uses policy $\pi$ ever after.

Let $V_{\pi^0}^b(b)$ be the expected total future payoff of a forager in belief state $b$ that takes the action of observing when in belief state $b$ and then uses policy $\pi$ ever after.

Let $V_{\pi^1}^b(b)$ be the expected total future payoff of a forager in belief state $b$ that takes the action of exploiting when in belief state $b$ and then uses policy $\pi$ ever after.

For this individual the possible equilibrium population distributions can be thought of as terminal states of a finite decision tree. Thus using dynamic programming we
can compute $V^{-1}_\pi(b)$, $V^0_\pi(b)$, and $V^1_\pi(b)$ for any reachable belief state.

With the values of each action in each belief state readily computable, we can apply a simple replicator dynamic [27] to the probabilities with which each action is chosen in each belief state, so that those actions with a higher expected value are chosen with higher probability than previously. The implementation of this process of iteratively applying stochastic dynamic programming and replicator dynamics can be found in appendix B.4. There is no guarantee that this replicator dynamic will converge on some fixed point in the candidate policy space, or that the point converged to is independent of the initial state of the replicator dynamic. However if this replicator dynamic does converge on a fixed point in the candidate policy space, this policy that is converged to will be an ES policy. For the all the parameter values tested the replicator dynamic did converge. Note that since the value of choosing each action is negatively frequency dependent, these ES policies have probabilities for choosing each action such that the expected future reward from choosing any of the actions chosen with positive probability is equal, in the belief states that foragers using this policy encounter.

To give a sense of what a policy actually looks like, consider a forager who finds themselves, along with the rest of the population, initially at a patch of quality 7. The forager does not know the quality of the other patch, but does know that this quality is distributed geometrically with parameter $\lambda = 0.1$. The forager also knows that the costs of observation and switching are determined by $c_o = 0.9$ and $c_s = 0.5$ and that this session will last for $T = 50$ rounds. With what probability should this forager exploit, observe or switch? According to table 2 on the first round the ES policy for the forager is to exploit with probability 0.25, observe with probability 0.23,
and switch to the other patch with probability 0.52. If the population finds its way to an equilibrium distribution as a result of these actions then every forager will just exploit until the end of the foraging session. If the population has not yet reached equilibrium after round two, then if this forager is one of the remaining forager’s at the starting patch it is again faced with a dilemma of whether to exploit, observe, or switch. Again table 2 gives the ES probabilities with which a forager remaining at the initial patch should choose each action.

<table>
<thead>
<tr>
<th>Round</th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>9</th>
<th>11</th>
<th>13</th>
<th>15</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploit</td>
<td>0.25</td>
<td>0.80</td>
<td>0.85</td>
<td>0.88</td>
<td>0.90</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
</tr>
<tr>
<td>Observe</td>
<td>0.23</td>
<td>0.05</td>
<td>0.04</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Switch</td>
<td>0.52</td>
<td>0.15</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round</th>
<th>19</th>
<th>21</th>
<th>23</th>
<th>25</th>
<th>27</th>
<th>29</th>
<th>31</th>
<th>33</th>
<th>35</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploit</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
</tr>
<tr>
<td>Observe</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Round</th>
<th>37</th>
<th>39</th>
<th>41</th>
<th>43</th>
<th>45</th>
<th>47</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploit</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.96</td>
<td>1.00</td>
</tr>
<tr>
<td>Observe</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Switch</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.04</td>
</tr>
</tbody>
</table>

### 4.3 Discussion

In the lone forager case there is always a single optimal action. In contrast in the many forager case, given the frequency dependent nature of the payoffs, the ES strategy always employs some blend of all three possible actions. The nature of a mixed ES strategy is such that the value of all actions used in the mixed strategy are equal. Thus in the many forager case the value of knowing “the quality of the other patch”
is totally dependent on how many other foragers also have this information. If the value of having this information is greater than the cost of obtaining it, then more foragers will observe. If the value of having this information is worth less than the cost of obtaining it then fewer foragers will observe, until a balance point is reached where the cost of obtaining the information equals the value that the information imparts. This leads us to the interesting, though unsurprising, conclusion that in a many forager scenario, with a population employing a mixed evolutionarily stable strategy which involves costly information gathering, the value of any information with an explicit cost will be exactly equal to that cost. This is in direct contrast to the lone forager optimization problem where information has a fixed value, and the optimal decision is based on comparing the fixed value of the information to the fixed cost of the information.
Chapter 5

Duck Problem 3

We are now ready to combine the problems of the previous two chapters so that, like the first pair of problems there is a variable environment and the foraging session is of uncertain duration, and like the second pair of problems there are explicit costs to switching, and observing. Unlike before we only consider the many foragers case as the single forager case is not significantly different from the single forager case in the chapter 3.

5.1 Many Foragers, Costly Switching, Costly Observation, Variable Terminal Time, Dynamic Environment

We now consider a problem with an arbitrarily large population of foragers, competing for the resources of two patches. Again a forager’s dilemma is one of determining how frequently to observe, and under which conditions it is worthwhile to switch from
one patch to the other. The proportion of the population that a forager shares a
patch with is a continuous variable, and so it is unfeasible to consider all possible
combinations of patch qualities and population distributions when determining the
ES frequencies of observation and switching thresholds. In problem 3.2 this difficulty
proved to not be a serious issue, since for each patch quality pair \((q_h, q_o)\) the policy
employed by the population defined a unique equilibrium population distribution.
In the current problem however, because there are explicit costs to switching from
one patch to the other just as in problem 4.2, there will not be a single equilibrium
population distribution for a given patch quality pair, \((q_h, q_o)\), but rather a continuum
of population distributions each with the property that informed foragers at both
patches have no incentive to switch patches. This was not an issue in problem 4.2,
where the patch qualities were static, since once the population was in an equilibrium
distribution it would stay in that distribution until the foraging session ended. In
contrast, in this current problem the patch qualities are dynamic, and so a population
will only be in an equilibrium distribution so long as the patch qualities do not change
(too much). Thus we will not be able to fully describe or compute an approximation
of an ES policy. We will however be able to discuss some of the properties of an ES
policy, and compute an approximation of an ES policy in certain specific situations.
Specifically we will find the approximate boundaries of the equilibrium population
distributions, for each patch quality pair, and approximate the ES rate of observation
when the population is in a distribution that is on the boundary of the equilibrium
population distributions.
Problem 5.1 Summary

Assumptions:

- Each round there is a patch quality change with probability $P_c$. Each patch is equally likely to be the patch that changes.
- If a patch quality changes, its new value is drawn from a geometric distribution with parameter $\lambda$.
- Each round, with probability $P_d$, the foraging session ends.
- Moving from one patch to another is costly, and that cost is proportional to the payoff at the patch being moved to.
- Observing one patch from the other is costly, though less so than moving between patches. This cost is in proportion to the payoff at the patch from which the observation is made.
- Foragers can make inferences about the other patch’s quality based on the movements of the population.
- The population is large enough that the proportion of the population employing a given action on a round is effectively deterministic, even if the action employed by any one individual is the stochastic result of a mixed strategy.
- A forager knows
  - The quality of its current patch.
  - The quality of the other patch if quality has not changed recently.
  - The current distribution of the population of foragers.
  - The parameters of the model, $P_c$, $P_d$, $\lambda$, $c_o$ and $c_s$.
- A forager does not know
  - The current quality of the other patch.

What we are looking for: The ES probability of observing the other patch, and when it is worthwhile to switch from one patch to the other, given the distribution of foragers and the current patch qualities.
5.1.1 POMDP formulation

As in the previous chapters the standard POMDP formulation can be augmented with the tools of evolutionary game theory to describe this problem. \( h \) (for here) and \( o \) (for other) are used to denote the patch that a forager has most recently foraged at, and the other patch respectively, relative to a specific forager, in the context where there is a specific forager, and are simply objective labels when there is no specific forager. Elements of the state space \( S \) consist of the current quality of the two patches, \( q_h \) and \( q_o \), and the proportion of the population at each patch \( d_h \) and \( d_o = 1 - d_h \). Then \( s \in S \) is of the form \( s = (q_h, q_o, d_h) \). The action space is \( A = \{-1, 0, 1\} \), where \( a = 1 \) indicates a forager exploiting at its current patch, \( a = 0 \) indicates observing the other patch while partially exploiting the current patch, and \( a = -1 \) indicates a forager migrating to the other patch.

The quality aspect of the state transitions, \( T((q_h^t, q_o^t), (q_h^{t+1}, q_o^{t+1})) \), are the same as in problem 3.1. The quality transition probabilities are as follows

\[
T((q_h^t, q_o^t), (q_h^{t+1}, q_o^{t+1})) = 1 - P_c + (P_c/2)\lambda(1 - \lambda)^{q_h - 1} + (P_c/2)\lambda(1 - \lambda)^{q_o - 1}
\]

if \( q_h = q_h' \) and \( q_o = q_o' \) (118)

\[
T((q_h^t, q_o^t), (q_h^{t+1}, q_o^{t+1})) = (P_c/2)\lambda(1 - \lambda)^{q_h' - 1}
\]

if \( q_h \neq q_h' \) and \( q_o = q_o' \) (119)

\[
T((q_h^t, q_o^t), (q_h^{t+1}, q_o^{t+1})) = (P_c/2)\lambda(1 - \lambda)^{q_o' - 1}
\]

if \( q_h = q_h' \) and \( q_o \neq q_o' \) (120)

Note that, as in problem 3.2 the patch quality transitions are totally independent of the actions of the foragers. In contrast the population distribution aspect of the state
is totally determined by the cumulative actions of the population.

\[ d_{h}^{t+1} = E[\text{proportion choosing } a = 0 \text{ and } a = 1 \text{ at the } h \text{ patch}] (d_{h}^{t}) + E[\text{proportion choosing } a = -1 \text{ at the } o \text{ patch}] (1 - d_{h}^{t}) \]

(121)

The reward function for a forager employing action \( a \) in state \( s^{t} = (q_{h}^{t}, q_{o}^{t}, d_{h}^{t}) \) then transitioning to state \( s^{t+1} = (q_{h}^{t+1}, q_{o}^{t+1}, d_{h}^{t+1}) \) is

\[
R_{1}(s^{t}, s^{t+1}) = q_{h}^{t+1} / d_{h}^{t+1}
\]

\[
R_{0}(s^{t}, s^{t+1}) = c_{o} \cdot q_{h}^{t+1} / d_{h}^{t+1}
\]

\[
R_{-1}(s^{t}, s^{t+1}) = c_{s} \cdot q_{h}^{t+1} / d_{h}^{t+1}
\]

(122)

The objective of an individual forager is to optimize the following

\[
\sum_{t=1}^{T} R_{a}(s^{t}, s^{t+1})
\]

(123)

where \( T \) is a geometrically distributed random variable with parameter \( P_{d} \). A forager does this by adopting a policy function, which gives the probability with which a forager should engage in each of the three possible actions, given the current belief state \( b \in B \) of the forager.

Each time step a forager receives information, which depends on the action taken by the forager that round. Every turn a forager receives information giving the quality of its current patch, the proportion of the population at the foragers current patch, and the action that the forager took. In addition if a forager has observed, \( a = 0 \), the forager also learns of the quality of the other patch. Thus the observation set is \( O = A \times Z^{+} \times [0,1] \times Z^{+} \cup \{0\} \). Since there is no noise on this observation, the new information a forager receives when it takes action \( a = 0 \) and transitions from state \( s^{t} = (q_{h}^{t}, q_{o}^{t}, d_{h}^{t}) \) to state \( s^{t+1} = (q_{h}^{t+1}, q_{o}^{t+1}, d_{h}^{t+1}) \) is \( o^{t+1} = (a^{t}, q_{h}^{t+1}, d_{h}^{t+1}, q_{o}^{t+1}) \) with probability one, and if it takes action \( a = 1 \) or \( a = -1 \) it receives the new information
o^t+1 = (a^t, q^t+1, d^t+1, 0), with probability one. Here the 0 in the last argument of o^t+1 indicates that the forager received no direct information about the quality of the other patch. The information available to the forager is its entire history of observations.

Assuming that the foragers have full knowledge of the dynamics of the patch qualities and that the foragers have full knowledge of the policy π employed by the population, then depending on the particular policy employed by the population, a forager may or may not be able to infer information about the quality of the other patch in the cases where it has not just observed. Thus, just as in problem 3.2, not only the dynamics of a forager’s belief state but also the very information that might inform this belief state is dependent on the policy adopted by the population.

5.1.2 The form of policies

**Definition 5.1.** Let $V_\pi(b)$ be the expected total future reward for a forager in belief state $b$ and employing policy $\pi$ in a population also employing policy $\pi$.

**Definition 5.2.** Let $V^1_\pi(b)$ be the expected total future reward for a forager in belief state $b$ that exploits on the next round, $a = 1$ and then follows policy $\pi$ ever after.

Let $V^0_\pi(b)$ be the expected total future reward for a forager in belief state $b$ that observes on the next round, $a = 0$ and then follows policy $\pi$ ever after, in a population using policy $\pi$.

Let $V^{-1}_\pi(b)$ be the expected total future reward for a forager in belief state $b$ that switches patches on the next round, $a = -1$ and then follows policy $\pi$ ever after.

In the case where the policy being used is an ES policy we use $V(b)$, $V^1(b)$, $V^0(b)$, $V^{-1}(b)$ instead of $V_\pi(b)$, $V^1_\pi(b)$, $V^0_\pi(b)$, $V^{-1}_\pi(b)$.
Definition 5.3. The population distribution is said to be in equilibrium, with respect to a policy $\pi$ employed by the population, if both of the following hold.

- The expected future value of all of the foragers at the same patch is approximately the same. Thus for any two belief states at the $h$ patch $b_h$ and $b'_h$ and for any two belief states at the $o$ patch, $b_o$ and $b'_o$

  \[
  V_\pi(b_h) \approx V_\pi(b'_h) \quad (124)
  \]

  \[
  V_\pi(b_o) \approx V_\pi(b'_o) \quad (125)
  \]

  There is an implicit assumption here that the policy employed by the population is such that the advantage or disadvantage resulting in having observed the quality of the other patch more recently is negligible, in the case that the other patch has not changed to something worth switching to.

- At each patch there is no incentive to switch patches, and some incentive to make observations that is

  \[
  V_\pi^1(b_h) = V_\pi^0(b_h) \approx V_\pi^1(b'_h) = V_\pi^0(b'_h) > V_\pi^{-1}(b_h) \approx V_\pi^{-1}(b'_h) \quad (126)
  \]

  \[
  V_\pi^1(b_o) = V_\pi^0(b_o) \approx V_\pi^1(b'_o) = V_\pi^0(b'_o) > V_\pi^{-1}(b_o) \approx V_\pi^{-1}(b'_o) \quad (127)
  \]

Definition 5.4. Let $d_{source}(q_h, q_o)$ be the maximum proportion of the population that can be at the $h$ patch when the patch qualities are $q_h$ and $q_o$ and where the system is in equilibrium with respect to the optimal policy.

Definition 5.5. Let $d_{sink}(q_h, q_o)$ be the minimum proportion of the population that can be at the $h$ patch when the patch qualities are $q_h$ and $q_o$ and where the system is in equilibrium with respect to the optimal policy. Note that $d_{sink}(q_h, q_o) = 1 - d_{source}(q_o, q_h)$. 
CHAPTER 5. DUCK PROBLEM 3

Definition 5.6. If the state of the system is in equilibrium, this equilibrium is called passive if \( d_{\text{sink}}(q_h, q_o) < d_h < d_{\text{source}}(q_h, q_o) \).

The equilibrium is called active if \( d_h = d_{\text{source}}(q_h, q_o) \) or if \( d_h = d_{\text{sink}}(q_h, q_o) \). In the case where \( d_h = d_{\text{source}}(q_h, q_o) \), the \( h \) patch is termed the source patch, and the \( o \) patch is termed the sink patch. In the case where \( d_h = d_{\text{sink}}(q_h, q_o) \) the \( h \) patch is termed the sink patch, and the \( o \) patch is termed the source patch.

How various patch changes create one equilibrium type or another is shown in figure 5.

Figure 5: Illustrations of how the foragers will respond to the quality of the source patch, \( h \), changing to \( q'_h \) or the quality of the sink patch \( o \) changing to \( q'_o \). \( q^*_h \) is the critical value \( q'_h \) must exceed in order to instigate migration from the former sink patch to the former source patch. \( q^*_h \) can be computed as the minimum \( q'_h \) satisfying \( d_h < 1 - d_{\text{source}}(q_o, q'_h) \). Similarly \( q^*_o \) is the critical value which \( q'_o \) must fall below in order to instigate migration from the former sink patch to the former source patch. \( q^*_o \) can be computed as the maximum \( q'_o \) satisfying \( d_h < 1 - d_{\text{source}}(q'_o, q_h) \).

Definition 5.7. Let \( O_{\text{source}}(q_h, q_o) \) be the probability with which foragers at a source patch with quality \( q_h \) takes action \( a = 0 \), when the population is in equilibrium, given
that the sink patch has quality $q_o$, such that the expected total future payoff from exploiting versus switching is the same, that is $V^1(\pi)(b_h) = V^0(\pi)(b_h')$, where $b_h$ and $b_h'$ are any of the belief states at the $h$ patch.

**Definition 5.8.** Let $O^i_{\text{source}}(q_h, q_o)$ and $S^i_{\text{source}}(q_h, q_o)$ be the probability with which foragers observe and switch respectively, in the following case. The forager is at a patch that was a source patch the last time there was an active equilibrium. This patch’s quality has not changed since that time. The forager knows that the other patch has changed to something worth switching to since all of the foragers that observed on the previous round have switched to the other patch. The forager did not observe this round and so does not know the quality of the other patch with certainty, only that it is greater than some lower bound for which there is incentive for some foragers to switch there. $i$ is the number of consecutive rounds where all of the foragers that have observed on the previous round have switched to the other patch on the next round. $q_h$ and $q_o$ were the qualities of the patches the last time there was an active equilibrium. $O^i_{\text{source}}$ and $S^i_{\text{source}}(q_h, q_o)$ are chosen so that one of the following holds.

1. $V^1(\pi)(b_h) = V^0(\pi)(b_h) = V^{-1}(\pi)(b_h)$
2. $V^1(\pi)(b_h) = V^0(\pi)(b_h) > V^{-1}(\pi)(b_h)$ and $S^i_{\text{source}} = 0$
3. $V^1(\pi)(b_h) = V^{-1}(\pi)(b_h) > V^0(\pi)(b_h)$ and $O^i_{\text{source}} = 0$
4. $V^0(\pi)(b_h) = V^{-1}(\pi)(b_h) > V^1(\pi)(b_h)$ and $S^i_{\text{source}} + O^i_{\text{source}} = 1$
5. $V^1(\pi)(b_h) > V^0(\pi)(b_h)$ and $V^1(\pi)(b_h) > V^{-1}(\pi)(b_h)$ and $S^i_{\text{source}} + O^i_{\text{source}} = 0$
6. $V^0(\pi)(b_h) > V^1(\pi)(b_h)$ and $V^0(\pi)(b_h) > V^{-1}(\pi)(b_h)$ and $O^i_{\text{source}} = 1$ and $S^i_{\text{source}} = 0$
7. $V^{-1}(\pi)(b_h) > V^2(\pi)(b_h)$ and $V^{-1}(\pi)(b_h) > V^0(\pi)(b_h)$ and $O^i_{\text{source}} = 0$ and $S^i_{\text{source}} = 1$
Here $b_h$ is the belief state of the foragers remaining at the $h$ patch that have not observed on the previous round, after $i$ rounds of all of the previous round’s observers leaving for the other patch. These functions are similar in purpose to the $S_i^i$ and $O_i^i$ functions of problem 4.2.

We let $O_{source}(q_h, q_o) = O_{source}^0(q_h, q_o)$ and $0 = S_{source}^0(q_h, q_o)$.

**Definition 5.9.** Let $O_{sink}(q_h, q_o)$ be the probability with which foragers at a sink patch with quality $q_h$ take action $a = 0$, given that the source patch has quality $q_o$, such that the expected total future payoff from exploiting versus switching is the same, that is $V_\pi^1(b_h) = V_\pi^0(b'_h)$, where $b_h$ and $b'_h$ are any of the belief states at the $h$ patch.

**Definition 5.10.** Let $O_{sink}^i(q_h, q_o)$ and $S_{sink}^i(q_h, q_o)$ be the probability with which foragers observe or switch respectively in the following case. The forager is at a patch that was a sink patch the last time there was an active equilibrium. This patch’s quality has not changed since that time. The forager knows that the other patch has changed to something worth switching to since all of the foragers that observed on the previous round have switched to the other patch. The forager did not observe this round and so does not know the quality of the other patch with certainty, only that it is greater than some lower bound for which there is incentive for some foragers to switch there. $i$ is the number of consecutive rounds where all of the foragers that have observed on the previous round have switched to the other patch on the next round. $q_h$ and $q_o$ are the former qualities of the patches during the most recent active equilibrium. $O_{sink}^i$ and $S_{sink}^i(q_h, q_o)$ are chosen so that one of the following holds.

1. $V_\pi^1(b_h) = V_\pi^0(b_h) = V_\pi^{-1}(b_h)$
2. $V_\pi^1(b_h) = V_\pi^0(b_h) > V_\pi^{-1}(b_h)$ and $S_{sink}^i = 0$
3. $V^i_\pi (b_h) = V^{-1}_\pi (b_h) > V^0_\pi (b_h)$ and $O^i_{sink} = 0$

4. $V^0_\pi (b_h) = V^{-1}_\pi (b_h) > V^1_\pi (b_h)$ and $S^i_{sink} + O^i_{sink} = 1$

5. $V^1_\pi (b_h) > V^0_\pi (b_h)$ and $V^1_\pi (b_h) > V^{-1}_\pi (b_h)$ and $S^i_{sink} + O^i_{sink} = 0$

6. $V^0_\pi (b_h) > V^1_\pi (b_h)$ and $V^0_\pi (b_h) > V^{-1}_\pi (b_h)$ and $O^i_{sink} = 1$ and $S^i_{sink} = 0$

7. $V^{-1}_\pi (b_h) > V^1_\pi (b_h)$ and $V^{-1}_\pi (b_h) > V^0_\pi (b_h)$ and $O^i_{sink} = 0$ and $S^i_{sink} = 1$

Here $b_h$ is the belief state of the foragers remaining at the $h$ patch that have not observed on the previous round, after $i$ rounds of all of the previous round’s observers leaving for the other patch.

**Definition 5.11.** Let $W_{\text{look}}(q_h, d_h, q_o, t, q^+_o, q^-_o)$ be the probability with which a forager observes and let $W_{\text{leap}}(q_h, d_h, q_o, t, q^+_o, q^-_o)$ be the probability with which a forager migrates to the other patch in the following case. The forager is at a patch that has recently changed quality, for the worse, so that now there is incentive for some foragers to leave this patch, given the lower bound on the quality of the other patch $q^-_o$. The forager’s patch has changed to quality $q_h$. There is currently $d_h$ of the total population at the $h$ patch. All of the foragers at this patch are able to place an upper bound $q^+_o$ on the quality of the other patch, a lower bound on the quality of the other patch $q^-_o$. This forager last observed the other patch, or noticed that informed foragers brought the system into an active equilibrium, $t$ rounds ago, and so the forager knows that the other patch had quality $q_o$, $t$ rounds ago. From this the forager can construct a belief distribution over the possible values of $q_o$. Note that different foragers at this patch will have different belief distributions over the possible qualities of the other patch, depending on when they last observed this other patch. Since the foragers are
assumed to know the policy of the population, foragers know the distribution of the most recent observation times of their patch mates. From this foragers are able to construct a belief distribution over the other foragers’ belief distributions of the value of $q_o$. $W_{\text{look}}(q_h, d_h, q_o, t, q_o^+, q_o^-)$ and $W_{\text{leap}}(q_h, d_h, q_o, t, q_o^+, q_o^-)$ are chosen so that one of the following holds

1. $V_1^1(b_h) = V_0^0(b_h) = V_{-1}^{-1}(b_h)$

2. $V_1^2(b_h) = V_0^0(b_h) > V_{-1}^{-1}(b_h)$ and $W_{\text{leap}} = 0$

3. $V_1^3(b_h) = V_{-1}^{-1}(b_h) > V_0^0(b_h)$ and $W_{\text{look}} = 0$

4. $V_0^0(b_h) = V_1^1(b_h) > V_{-1}^{-1}(b_h)$ and $W_{\text{leap}} + W_{\text{look}} = 1$

5. $V_1^4(b_h) > V_0^0(b_h)$ and $V_1^1(b_h) > V_{-1}^{-1}(b_h)$ and $W_{\text{leap}} + W_{\text{look}} = 0$

6. $V_0^0(b_h) > V_1^1(b_h)$ and $V_0^0(b_h) > V_{-1}^{-1}(b_h)$ and $W_{\text{look}} = 1$ and $W_{\text{leap}} = 0$

7. $V_{-1}^{-1}(b_h) > V_1^1(b_h)$ and $V_{-1}^{-1}(b_h) > V_0^0(b_h)$ and $W_{\text{look}} = 0$ and $W_{\text{leap}} = 1$

Thus a policy $\pi$ can be partially described in terms of the functions, $d_{\text{source}}$, $d_{\text{sink}}$, $O_{\text{source}}^i$, $O_{\text{sink}}^i$, $S_{\text{source}}^i$, $S_{\text{sink}}^i$, $W_{\text{look}}$ and $W_{\text{leap}}$.

The cases that follow are those that can be totally described by these functions.

We assume that the forager’s belief state, that is the information used by the policy function to determine the probability of using each action, consists of the following.

1. The quality of it’s current patch, $q_h$.

2. The quality of the other patch the last time this was known with certainty, $\hat{q}_o$.

3. The number of rounds since the forager last knew the quality of the other patch with certainty, $t$. 
4. An inferred lower bound on the quality of the other patch, \( q_o^- \), given either that none of the foragers at the other patch have switched to the forager’s patch or that all of the forager’s patch mates that observed on the previous round have switched to the other patch.

5. An inferred upper bound on the quality of the other patch given that none of the foragers at the forager’s current patch have switched to the other patch \( q_o^+ \).

Note that \( q_o^+ = +\infty \) in the case where all of the foragers that observed the other patch on the previous round switch to the other patch on the current round.

6. The patch qualities of both patches the last time the population was in an active equilibrium distribution \( q^{e}_h \) and \( q^{e}_o \).

7. The current distribution of the foragers, \( d_h \).

8. Whether or not its current patch was most recently a sink patch or a source patch. This can often inferred from whether or not \( d_h = d_{\text{source}}(q^{e}_h, q^{e}_o) \) or if \( d_h = d_{\text{sink}}(q^{e}_h, q^{e}_o) \), although in some cases a forager just has to remember this.

Case A. The forager’s patch was a source patch the last time there was an active equilibrium and the forager’s patch has not changed quality since that time, that is \( q_h = q^{e}_h \) and \( d_h = d_{\text{source}}(q^{e}_h, q^{e}_o) \). So far as the forager knows the other patch has not changed to anything worth switching to, or from. That is if the other patch has changed quality to \( q'_o \), the following is satisfied, \( d_{\text{sink}}(q_h, q'_o) \leq d_h \leq d_{\text{source}}(q_h, q'_o) \), which places bounds on \( q'_o \). In this case the forager will observe with probability \( O_{\text{source}}(q^{e}_h, q^{e}_o) \). If the forager does not observe then it will exploit. The forager will never switch patches in this case.
Case B. The forager’s patch was a source patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time, that is \( q_h = q_h^o \) and \( d_h = d_{source}(q_h^o, q_o^e) \). For compactness we let \( O = O_{source}(q_h^o, q_o^e) \), \( d = d_{source}(q_h^o, q_o^e) \), \( O' = O_{source}(q_h, q_o^e) \) and \( d' = d_{source}(q_h, q_o^o) \). The forager knows that the other patch has just changed to \( q_o^o \), and is worth switching to since \( d_{source}(q_h, q_o^o) < d_h \). The forager knows this because this forager was one of \( O \cdot d \) foragers at the patch that observed this round. In this case the forager will migrate from this patch with probability \( P_m \), such that if possible the population distribution will be in active equilibrium on the next round, that is

\[
d - d' = d \cdot O \cdot P_m
\]

and so solving for \( P_m \) we have

\[
P_m = \min \left[ 1, \frac{d - d'}{d \cdot O} \right]
\]

(128)

(129)

If the forager stays at this patch, which it will do with probability \( 1 - P_m \) then the forager will observe with probability \( P_o \) so that if possible

\[
d' \cdot O' = d \cdot (1 - O) \cdot O + d \cdot (1 - O) \cdot (1 - P_m) \cdot (P_o)
\]

(130)

and so solving for \( P_o \) we have,

\[
P_o = \max \left[ 0, \min \left[ 1, \frac{d' \cdot O' - d \cdot (1 - O) \cdot O}{d \cdot (1 - O) \cdot (1 - P_m)} \right] \right]
\]

(131)

Thus a forager in this case takes action \( a = -1 \) with probability \( P_m \), takes action \( a = 0 \) with probability \( (1 - P_m)P_o \) and takes action \( a = 1 \) with probability \( (1 - P_m)(1 - P_o) \).

Case C. The forager’s patch was a source patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. For
compactness we let \( O^i = O^i_{source}(q_h^c, q_o^c) \), \( S^i = S^i_{source}(q_h^c, q_o^c) \), \( d = d_{source}(q_h^c, q_o^c) \), \( d' = d_{source}(q_h^c, q_o^c) \), and \( O' = O_{source}(q_h, q_o^c) \). Further we let

\[
d^n = d \cdot \prod_{i=0}^{n} (1 - S^i - O^i)
\]
denote the proportion of the total population remaining at the former source patch, that have not observed the other patch during the \( n \) rounds since the quality of the former sink patch improved enough to cause migration, given that the population is not yet in an equilibrium distribution. The forager knows that the other patch has changed to something worth switching to on a previous round because all of the foragers that observed on previous rounds have switched to the other patch. The forager is one of the \( O^n \cdot d^{n-1} \) foragers remaining at the patch that observed the quality of the other patch on this round. In this case, the other foragers at the patch are also aware that the quality of the other patch has improved enough that there is incentive to switch there, and so \( S^{n+1} \cdot d^n \) foragers will blindly switch to the other patch on the next round. This informed forager will take this into account, and migrate to the other patch with probability \( P_m \), so that if possible the population distribution is in an active equilibrium on the next round, that is

\[
d^n \cdot (1 - S^{n+1}) - d' = d^{n-1} \cdot O^n \cdot P_m
\]

and so solving for \( P_m \) we have,

\[
P_m = \max \left[ 0, \min \left[ 1, \frac{d^n \cdot (1 - S^{n+1}) - d'}{d^{n-1} \cdot O^n} \right] \right]
\]

If some, but not all of these informed foragers remains behind, they will observe
with probability $P_o$, so that if possible
\[
O' \cdot d' = O^{n+1} \cdot d^n + O^n \cdot d^{n-1} \cdot (1 - P_m) \cdot P_o
\] (134)

Solving for $P_o$ we have that
\[
P_o = \max \left[ 0, \min \left[ 1, \frac{O' \cdot d' - O^{n+1} \cdot d^n}{O^n \cdot d^{n-1} \cdot (1 - P_m)} \right] \right]
\] (135)

The probability with which these informed foragers should observe in the case where $P_m = 0$, cannot be described using just the functions $d_{source}$, $d_{sink}$, $O^i_{source}$, $O^i_{sink}$, $S^i_{source}$, $S^i_{sink}$, $W_{look}$ and $W_{leap}$.

Case D. The forager’s patch was a source patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. For compactness we let $O^i = O^i_{source}(q^c_h, q^c_o)$, $S^i = S^i_{source}(q^c_h, q^c_o)$, $d = d_{source}(q^c_h, q^c_o)$, and $d' = d_{source}(q_{h}', q_{o}')$. Further we let
\[
d^n = d \cdot \prod_{i=0}^{n} (1 - S^i - O^i)
\]
denote the proportion of the total population remaining at the former source patch, that have not observed the other patch during the $n$ rounds since the quality of the former sink patch improved enough to cause migration, given that the population is not yet in an equilibrium distribution. The forager knows that the other patch has changed to something worth switching to on a previous round because all of the foragers that observed on previous rounds have switched to the other patch. The forager is not one of the $O^n \cdot d^{n-1}$ foragers remaining at the former source patch that has observed the quality of the other patch on this current round. In this case, the forager is aware that the quality of the other patch has improved enough that there is incentive to switch there, but does not know the quality of the other patch precisely and so the forager will
observe with probability $O^{n+1}$, switch to the other patch with probability $S^{n+1}$ and exploit with probability $1 - O^{n+1} - S^{n+1}$.

Case E. The forager’s patch was a source patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. A forager has observed that the other patch just worsened to $q'_o$, and so knows that $(1 - d) \cdot W_{\text{leap}}$ foragers will be switching to its patch, and that this switching will bring about an active equilibrium. For compactness we let $O = O_{\text{source}}(q_h, q'_o)$, $d = d_{\text{source}}(q_h, q'_o)$, $O' = O_{\text{sink}}(q_h, q'_o)$ and $d' = d_{\text{sink}}(q_h, q'_o)$. The forager will observe with probability $P_o$ such that if possible

$$O' \cdot d' = d \cdot (1 - O) \cdot O + d \cdot O \cdot P_o$$

Solving for $P_o$ we have that

$$P_o = \max \left[ 0, \min \left[ 1, \frac{O' \cdot d' - d \cdot (1 - O) \cdot O}{d \cdot O} \right] \right]$$

Case F through Case J are the sink patch analogues of Case A through Case E. They are described in detail in appendix A.4.

Regardless of whether a forager is at a source patch or a sink patch, or a patch that is neither source nor sink, the $W_{\text{look}}$ and $W_{\text{leap}}$ functions inform what a forager does in the following case.

Case K. The forager’s patch changes from $q_h$ to $q'_h$ with $q'_h < q_h$ so that now there is incentive for some of the foragers at this patch to leave. In this case the forager observes with probability $W_{\text{look}}(q_h, q'_h, d_h, q_o, t, q_o^+, q_o^-)$, migrates to the other patch with probability $W_{\text{leap}}(q_h, q'_h, d_h, q_o, t, q_o^+, q_o^-)$, and exploits with probability $1 - W_{\text{look}}(q_h, q'_h, d_h, q_o, t, q_o^+, q_o^-) - W_{\text{leap}}(q'_h, d_h, q_o, t, q_o^+, q_o^-)$. 
We now turn our attention to the cases where the policy cannot be described using only functions $d_{source}, d_{sink}, O'_{source}, O'_{sink}, W_{look}$ and $W_{leap}$. Consider the two following cases.

Case L. The forager’s patch was a source patch the last time there was an active equilibrium, and the forager’s current patch has changed on a previous round from $q^e_h$ to $q_h$, with $q^e_h < q_h$, but this improvement is not large enough for foragers at the other patch to switch to the forager’s current patch, so that $d_h$ is still equal to $d_{source}(q^e_h, q^e_o)$, and to the best of this forager’s knowledge the quality of the other patch has not changed to anything worth switching to or from. Thus the system is now, if it was not already in a passive equilibrium. We not attempt to describe the policy in detail. We only note that the in such a situation the policy should be such that there is no switching to the other patch, and there is some observing and that this probability of observing is bounded above by $O_{source}(q^e_h, q^e_o)$ and bounded below by $O_{sink}(q^e_h, q^e_o)$.

Case M. The forager’s patch was a sink patch the last time there was an active equilibrium, and the forager’s current patch has changed on a previous round from $q^e_h$ to $q_h$, with $q_h < q^e_h$, but this worsening is not large enough for foragers at this patch to switch to the other patch, so that $d_h$ is still equal to $d_{sink}(q^e_h, q^e_o)$, and to the best of this forager’s knowledge the quality of the other patch has not changed to anything worth switching to or from. We again do not attempt to describe the policy in detail. We only note that in such a situation the policy should be such that there is no switching to the other patch, and there is some observing and that this probability of observing is bounded above by $O_{source}(q^e_h, q^e_o)$ and bounded below by $O_{sink}(q^e_h, q^e_o)$.
Case L and Case M make clear how the population can transition into the passive equilibrium analogues of cases Case F and Case A. Case B through to Case J all deal with how foragers respond to changes that are likely to eventually bring about an active equilibrium, given that the foragers responding to these changes are in an active equilibrium (or at least observing with the same frequencies that they would use in an active equilibrium). To complete this policy description would require considering the passive equilibrium analogues for cases Case B through to Case J, which we do not do here.

5.1.3 Approximating an ES policy

While we do not attempt a complete characterization of the form of the ES policy, we do compute approximations of the functions $d_{\text{source}}$, $d_{\text{sink}}$, $O_{\text{source}}$ and $O_{\text{sink}}$, used by an ES policy.

Difference in total future expected payoff at Source versus Sink patch

Given functions $d_{\text{source}}$, $d_{\text{sink}}$, $O_{\text{source}}$ and $O_{\text{sink}}$ an approximation of the difference, $D(q_h, q_o)$, in expected total future reward resulting from being a forager at the sink patch versus being a forager at the source patch, in an active equilibrium, where the sink patch has quality $q_h$ and the source patch has quality $q_o$, can be computed by conditioning on the type of the first change (patch quality change or session ending), and making some assumptions about the behaviour produced by an ES policy. For compactness we let $d_h = d_{\text{sink}}(q_h, q_o)$, $d_o = d_{\text{source}}(q_o, q_h)$, $O_o = O_{\text{source}}(q_o, q_h)$, $O_h = O_{\text{sink}}(q_h, q_o)$, and $D = D(q_h, q_o)$.

On average there will be $\delta/(1 - \delta)$ with $\delta = (1 - P_c)(1 - P_d)$, rounds prior to the
first change occurring. A forager at the sink patch receives an average payoff of

$$r_h = (1 - O_h \cdot c_o) \left( \frac{q_h}{d_h} \right)$$

(138)
on each of these rounds, and a forager at the source patch receives an average payoff of

$$r_o = (1 - O_o \cdot c_o) \left( \frac{q_o}{d_o} \right)$$

(139)
which allows us to compute the contribution to $D$ from the rounds prior to the first change, as

$$D_{prior} = (r_h - r_o) \frac{\delta}{1 - \delta}$$

(140)

Once a change occurs with probability $P_d/(1-\delta)$ this change is the foraging session ending, and there is no further contribution to $D$. With probability $P_a = 1 - P_d/(1-\delta)$ this change is a patch quality changing, and there are further contributions to $D$.

Thus on the round that the patch quality changes the expected contribution to $D$ is

$$D_{during} = P_a \left[ \frac{(q_h + 1/\lambda)(1 - O_h \cdot c_o)}{d_h} - \frac{(q_o + 1/\lambda)(1 - O_o \cdot c_o)}{d_o} \right]$$

(141)

If a patch quality change occurs, it is of one of the six following types.

change type 1. The $h$ patch improves to quality $q'_h$ so that $d_{source}(q_o, q'_h) < d_o$, and so some foragers will be switching to the $h$ patch.

change type 2. The $o$ patch improves to quality $q'_o$ so that $d_{source}(q_h, q'_o) < d_h$, and so some foragers will be switching to the $o$ patch.

change type 3. The $h$ patch worsens to quality $q'_h$ so that $d_{sink}(q_o, q'_h) \leq d_o \leq d_{source}(q_o, q'_h)$,

and so even though there was a patch quality change no foragers will be switching from one patch to the other.
change type 4. The \( o \) patch improves to quality \( q'_o \) so that \( d_{\text{sink}}(q_h, q'_o) \leq d_h \leq d_{\text{source}}(q_h, q'_o) \), and so even though there was a patch quality change no foragers will be switching from one patch to the other.

change type 5. The \( h \) patch worsens to quality \( q'_h \) so that \( d_{\text{source}}(q'_h, q_o) < d_h \), and so some foragers will be switching to the \( o \) patch.

change type 6. The \( o \) patch worsens to quality \( q'_o \) so that \( d_{\text{source}}(q'_o, q_h) < d_o \), and so some foragers will be switching to the \( h \) patch.

If change type 1 occurs, to compute the approximate contribution to \( D \), we condition on the possible values of \( q'_h \) and calculate the probability of it taking, 1, 2 or more rounds for enough of the foragers at the \( o \) patch to switch to the \( h \) patch and bring about equilibrium.

\( O_o \) of the \( d_o \) foragers at the \( o \) patch will become aware of the change at the \( h \) patch as soon as it happens. Thus \( O_o \) of the remaining \((1 - O_o) \cdot d_o\) foragers at the \( o \) patch will be aware of the precise nature of the change one round after it happens, where as \( 1 - O_o \) of the remaining \((1 - O_o) \cdot d_o\) foragers at the \( o \) patch will be aware that the other patch has improved enough that there is incentive for some foragers to switch there, but will not be aware of the precise nature of the change at the other patch. We assume, for analytic ease, that \( S^1_{\text{source}} = S^1_{\text{sink}} = 0 \).

Thus if \( q'_h \) is such that

\[
d_o - d_{\text{source}}(q_o, q'_h) < O_o \cdot d_o
\] (142)

then equilibrium will be achieved on the round immediately following the change. In this case we know that the foragers that switched to the \( h \) patch have the same expected total future reward as those that stayed at the \( o \) patch, and those that
switch to the \( h \) patch have approximately the same total expected future reward as those currently at the \( h \) patch, less the cost of switching to the \( h \) patch. Then the approximate contribution to \( D \) in this case, weighted by the probability of this case is

\[
D_{after} = \frac{P_a}{2} \cdot P_{1,1} \cdot \delta (1 - c_s) E \left[ \frac{q'_h}{d_{sink}(q'_h, q_o)} \right]
\]

(143)

Where the expectation is taken over the possible value of \( q'_h \) conditional on \( q'_h > q_h \) and inequality 142, and \( P_{1,1} \) is the probability that \( q'_h \) satisfies these conditions. Note that we have discounted this contribution by \( \delta \), the probability of there not being another patch quality change immediately following the first patch quality change. Thus we are assuming that if there is a second change immediately following the first change, the expected contribution to \( D \) during and after this second change is small.

If \( q'_h \) is such that

\[
d_o - d_{source}(q_o, q'_h) \geq O_o \cdot d_o,
\]

(144)

but

\[
d_o - d_{source}(q_o, q'_h) - O_o \cdot d_o < 0_o (1 - O_o) d_o,
\]

(145)

then equilibrium will be achieved two rounds after the patch quality changes. In this case the foragers already at the \( h \) patch have an advantage over those foragers that switch to the \( h \) patch on the round immediately after the change, and an advantage over those foragers that switch to the \( h \) patch two rounds after the change. Reasoning as before the approximate contribution to \( D \) in this case, weighted by the probability
of this case is

\[ D_{after} = P_a/2 \cdot P_{1.2} \left( (1 - O_o) \left( \delta^2 (1 - cs) E \left[ \frac{q_h'}{d_{sink}(q_h', q_o)} \right] 
+ \delta \left( \frac{E[q_h']}{d_h + O_o \cdot d_o} - \frac{q_o}{d_o \cdot (1 - O_o)} \right) \right) 
+ O_o \left( \delta (1 - cs) \frac{E[q_h']}{d_h + O_o \cdot d_o} \right) \right) \]  

(146)

Where the expectations are taken over \( q_h' \) given that \( q_h' \) satisfies inequalities 144 and 145, and \( P_{1.2} \) is the probability that \( q_h' \) satisfies these conditions.

If

\[ d_o - d_{source}(q_o, q_h) - O_o \cdot d_o \geq 0_o(1 - O_o)d_o, \]  

(147)

we assume that \( O_{source}(q_o, q_h) \) and \( S_{source}(q_h, q_o) \) is such that the probability of it ever taking more than three rounds for enough of the foragers at the \( o \) patch to switch to the \( h \) patch and bring about equilibrium is negligible. In this case the approximate contribution to \( D \) weighted by the probability of this case is

\[ D_{after} = P_a/2 \cdot P_{1.3} \left( (1 - O_o)^2 \left( \delta^3 (1 - cs) E \left[ \frac{q_h'}{d_{sink}(q_h', q_o)} \right] 
+ \delta^2 \left( \frac{E[q_h']}{d_h + d_o \cdot (1 - (1 - O_o)^2)} - \frac{q_o}{d_o \cdot (1 - O_o)^2} \right) \right) 
+ \delta \left( \frac{E[q_h']}{d_h + O_o \cdot d_o} - \frac{q_o}{d_o \cdot (1 - O_o)} \right) \right) 
+ O_o \cdot (1 - O_o) \left( \delta^2 (1 - cs) \frac{E[q_h']}{d_h + O_o \cdot d_o + O_o \cdot (1 - O_o) \cdot d_o} \right) 
+ \delta \left( \frac{E[q_h']}{d_h + O_o \cdot d_o} - \frac{q_o}{d_o \cdot (1 - O_o)} \right) \right) 
+ O_o \left( \delta (1 - cs) \frac{E[q_h']}{d_h + O_o \cdot d_o} \right) \]  

(148)

Where the expectations are taken over \( q_h' \) given that \( q_h' \) satisfies inequality 147, and \( P_{1.3} \) is the probability that \( q_h' \) satisfies these conditions.
If change type 2 occurs, to compute the approximate contribution to $D$, we condition on the possible values of $q'_o$ and calculate the probability of it taking, 1, 2 or more rounds for enough of the foragers at the $h$ patch to switch to the $o$ patch and bring about equilibrium.

$O_h$ of the $d_h$ foragers at the $h$ patch will become aware of the change at the $o$ patch as soon as it happens. $O_h$ of the remaining $(1 - O_h)d_h$ foragers at the $h$ patch will become aware of the change one round after it happens.

Thus if

$$d_h - d_{\text{source}}(q_h, q'_o) < O_h \cdot d_h$$  \hspace{1cm} (149)$$

then equilibrium will be achieved on the round immediately following the change. In this case we know that the foragers that switched to the $o$ patch have the same expected total future reward as those that stayed at the $h$ patch, and those that switch to the $o$ patch have approximately the same total expected future reward as those currently at the $o$ patch, less the cost of switching to the $o$ patch. Then the approximate contribution to $D$ in this case, weighted by the probability of this case is

$$D_{\text{after}} = P_a/2 \cdot P_{2.1} \cdot \delta(c_s - 1)E \left[ \frac{q'_o}{d_{\text{sink}}(q'_o, q_h)} \right]$$  \hspace{1cm} (150)$$

Where the expectation is taken over $q'_o$ given $q'_o$ satisfies inequality 149 and that $q'_o \geq q'_o^*$, where $q'_o^*$ is the minimum value that the $o$ patch must improve to, for foragers from the $h$ patch to migrate to the $o$ patch. $P_{2.1}$ is the probability with which $q'_h$ satisfies these conditions. Again we discount this contribution by $\delta$, since we assume, as before, that if there is a second change immediately following the first change, the expected contribution to $D$ during and after this second change is small.
If
\[ d_h - d_{source}(q_h, q'_o) \geq O_h \cdot d_h, \tag{151} \]
but
\[ d_h - d_{source}(q_h, q'_o) - O_h \cdot d_h < 0_h(1 - O_h)d_h, \tag{152} \]
then equilibrium will be achieved two rounds after the patch quality changes. In this case the foragers already at the \( o \) patch have an advantage over those foragers that switch to the \( o \) patch on the round immediately after the change, and an advantage over those foragers that switch to the \( o \) patch two rounds after the change. Reasoning as before the approximate contribution to \( D \) in this case, weighted by the probability of this case is
\[
D_{after} = P_a/2 \cdot P_{2,2} \left( (1 - O_h) \left( \delta^2 (c_s - 1) E \left[ \frac{q'_o}{d_{sink}(q'_o, q_h)} \right] + \delta \left( \frac{q_h}{d_h \cdot (1 - O_h)} - \frac{E[q'_o]}{d_o + O_h \cdot d_h} \right) \right) + O_h \left( \delta (c_s - 1) \frac{E[q'_o]}{d_o + O_h \cdot d_h} \right) \right) \tag{153}
\]
Where the expectations are taken over \( q'_o \) given that \( q'_o \) is such that inequalities 151 and 152 are satisfied, and where \( P_{2,2} \) is the probability with which \( q'_o \) satisfies these conditions.

If
\[ d_h - d_{source}(q_h, q'_o) - O_h \cdot d_h \geq 0_h(1 - O_h)d_h, \tag{154} \]
we assume that \( O_{sink}^i(q_h, q_o) \) and \( S_{sink}^i(q_h, q_o) \) are such that the probability of it ever taking more than three rounds for enough of the foragers at the \( h \) patch to switch to the \( o \) patch and bring about equilibrium is negligible. In this case the approximate
contribution to $D$ weighted by the probability of this case is

$$D_{after} = P_a/2 \cdot P_{2.3} \left( (1 - O_h)^2 \left( \delta^3 (cs - 1) E \left[ \frac{q'_o}{d_{sink}(q'_o, q_h)} \right] \right. \right.$$  
$$+ \delta^2 \left( \frac{q_h}{d_h \cdot (1 - O_h)^2} - \frac{E[q'_o]}{d_o + d_h \cdot (1 - (1 - O_h)^2)} \right) \right.$$  
$$+ \delta \left( \frac{q_h}{d_h \cdot (1 - O_h)} - \frac{E[q'_o]}{d_o + O_h \cdot d_h} \right) \right)$$  
$$+ O_h \cdot (1 - O_h) \left[ \delta^2 (cs - 1) \frac{E[q'_o]}{d_o + O_h \cdot d_h + O_h \cdot (1 - O_h) \cdot d_h} \right.$$  
$$+ \delta \left( \frac{q_h}{d_h \cdot (1 - O_h)} - \frac{E[q'_o]}{d_o + O_h \cdot d_h} \right) \right)$$  
$$+ O_o \left( \delta (cs - 1) \frac{E[q'_o]}{d_o + O_h \cdot d_h} \right) \right). \tag{155}$$

Where the expectations are taken over $q'_o$ given that $q'_o$ is such that inequality 154 is satisfied, and $P_{2.3}$ is the probability with which $q'_o$ satisfies this condition.

If change type 3 occurs where the former sink patch, $h$, worsens to quality $q'_h$ so that

$$d_{sink}(q_o, q'_h) \leq d_o \leq d_{source}(q_o, q'_h), \tag{156}$$

then even though there was a patch quality change no foragers will be switching from one patch to the other. An approximation of the expected contribution to $D$ after the first change in this case weighted by the probability of this case is

$$D_{after} = P_a/2 \cdot P_3 \cdot \frac{\delta}{1 - \delta} \left( (1 - O_h \cdot c_o) \frac{E[q'_h]}{d_h} - (1 - O_o \cdot c_o) \frac{q_o}{d_o} \right) \tag{157}$$

Where the expectation over $q'_h$ is conditional on $q'_h$ being such that inequality 156 is satisfied, and where $P_3$ is the probability that $q'_h$ satisfies this condition. This is an approximation since the true probability of observing at the changed $h$ patch is not $O_h$, and because this expression only takes into account the contribution to $D$ after the first change but prior to the second change. To improve this approximation would
require conditioning on the type of the second change which we do not do here.

If change type 4 occurs, where the former source patch, $o$, improves to quality $q'_o$ so that
\[
d_{sink}(q_h, q'_o) \leq d_h \leq d_{source}(q_h, q'_o),
\]
and so even though there was a patch quality change no foragers will be switching from one patch to the other. An approximation of the expected contribution to $D$ after the first change in this case weighted by the probability of this case is
\[
D_{after} = P_a/2 \cdot P_4 \cdot \frac{\delta}{1 - \delta} \left(1 - O_h \cdot c_o\right) \frac{q_h}{d_h} - \left(1 - O_o \cdot c_o\right) \frac{E[q'_o]}{d_o}
\]
(159)

Where the expectation over $q'_o$ is conditional on $q'_o$ being such that inequality 158 is satisfied, and where $P_4$ is the probability that $q'_o$ satisfies this condition. This is an approximation since the true probability of observing at the changed $o$ patch is not $O_o$, and because this expression only takes into account the contribution to $D$ after the first change but prior to the second change. To improve this approximation would require conditioning on the type of the second change, which we do not do here.

If change type 5 occurs, the former sink patch, $h$, worsens to quality $q'_h$ so that
\[
d_{source}(q'_h, q_o) < d_h,
\]
and so some of the foragers at the $h$ patch should switch to the $o$ patch. We assume that the functions $W_{look}$ and $W_{leap}$, are such that the total expected future payoffs of the foragers is approximately the same as if an active equilibrium is achieved on the round immediately following the change. Thus the approximate contribution to $D$ after the first change, in this case, is simply the cost of switching patches, that is
\[
D_{after} = P_a/2 \cdot P_5 \cdot \delta(c_s - 1) E_{q_h} \left(\frac{q_o}{d_{sink}(q_o, q'_h)}\right)
\]
(161)

Where the expectation is taken over $q'_h$ given that $q'_h$ is such that inequality 160 is
satisfied, and where $P_5$ is the probability that $q'_h$ satisfies this condition. As usual this contribution is weighted by the probability of the case.

If change type 6 occurs, the former source patch, $o$, worsens to quality $q'_o$ so that

$$d_{source}(q_o, q'_h) < d_o,$$

and so some of the foragers at the $o$ patch should switch to the $h$ patch. We assume that the functions $W_{look}$ and $W_{leap}$, are such that the total expected future payoffs of the foragers is approximately the same as if an active equilibrium is achieved on the round immediately following the change. Thus the approximate contribution to $D$ after the first change, in this case, is simply the cost of switching patches, that is

$$D_{after} = P_a/2 \cdot P_6 \cdot \delta(1 - c_s)E\left[\frac{q_h}{d_{sink}(q_h, q'_o)}\right]$$

Where the expectation is taken over $q'_o$ given that $q'_o$ is such that inequality 162 is satisfied, and where $P_6$ is the probability that $q'_o$ satisfies this condition. As usual this contribution is weighted by the probability of the case.

A computer program for computing an approximation of, $D(q_h, q_o)$, the difference in total future expected payoff between a forager at a sink patch of quality $q_h$ versus a source patch of quality $q_o$, given the functions $d_{source}$, $d_{sink}$, $O_{source}$ and $O_{sink}$, as outlined above, can be found in appendix B.5.

**Difference in total future expected payoff when Observing versus Not**

Just as in problem 4.2 (recall lemma 4.13) an observation is only of value if it results in the discovery that the other patch is of a high enough quality, that all of the foragers that observed this round have incentive to switch to the other patch. Now a forager can tell from the fact that other foragers at its patch have not switched to the other patch, that if the other patch has changed to something worth switching to,
this has happened on the current round. Thus a forager in this problem will observe
in the hopes that, either it will observe a large improvement in the quality of the
other patch on the very round that change occurs, or that it will observe a very large
improvement in the quality of the other patch, the round after the change occurs,
when all of the foragers that observed on the previous round have switched to the
other patch.

Thus for there to be any gain from observation for a forager in an equilibrium
situation, the other patch must change quality on the very round that the forager
observes, which happens with probability

\[ (1 - P'_c) \cdot P_c / 2 \]  

or in some cases on a forager may also gain from observation in the case where the
other patch has changed quality on the round prior to when the forager observes
which happens with probability

\[ P'_c = \frac{P_c}{2 - P_c(1 - \lambda(1 - \lambda)^{q_h-1})} \]  

As in previous problems, \( P'_c \) is the probability that the other patch has changed on
the previous round, given that a forager’s patch did not perceptively change on the
previous round.

Note that when computing this approximation we will not consider the case where
there is a patch quality change immediately following another patch quality change.

For compactness we let \( d_h = d_{source}(q_h, q_o), \ d_o = d_{sink}(q_o, q_h), \ d'_o = d_{sink}(q'_o, q_h), \ O^i = O^i_{source}(q_h, q_o) \) and \( S^i = S^i_{source}(q_h, q_o) \). Making the same assumptions about the
policy employed by the population as we did when computing an approximation of
\( D \), when the other patch’s quality, \( q_o \), changes to \( q'_o \) one of three situations will be
the case.
Case 1. The new quality of the other patch, is such that fewer than all, perhaps even none, of the foragers that observed this round have incentive to switch to the other patch, that is

\[ d_h - d'_h < O_h \cdot d_h \]  

(166)

In this case not all of the foragers that are aware of the change will switch to the other patch, and the foragers that switch to the other patch will have the same expected total future reward as those that stay regardless of whether those that stayed observed the round previously, and so those that observed will have gained nothing.

Case 2. The new quality of the other patch, \( q'_o \), is such that all of the foragers that observed this round, but not all of the foragers that observe next round, have incentive to switch to the other patch, that is

\[ O_h \cdot d_h < d_h - d'_h < O_h \cdot (1 - O_h) \cdot d_h + O_h \cdot d_h \]  

(167)

In this case the foragers that first notice this change will gain approximately

\[ \delta^2 (1 - c_s) \frac{d'_o}{d'_o} + \delta \left( \frac{c_s \cdot q'_o}{d_o + O_h \cdot d_h} - \frac{q_h}{(1 - O_h) \cdot d_h} \right) \]  

(168)

as compared to the foragers that did not observe.

Case 3. The new quality of the other patch, \( q'_o \), is such that all of the foragers that observed this round, and all of the foragers that observe next round, have incentive to switch to the other patch, that is

\[ O_h \cdot (1 - O_h) \cdot d_h + O_h \cdot d_h < d_h - d'_h \]  

(169)

As before we assume that \( S^i \) and \( O^i \) are such that the probability of it ever taking more than three rounds after a change occurs for the forager’s at the
unchanged patch to become aware of this change to bring about equilibrium is negligible.

In this case the foragers that first notice this change will gain approximately

\[ (1 - O_h) \left( \delta^3 (1 - cs) \frac{q'_o}{d'_o} + \delta^2 \left( \frac{q'_o}{d_o + (1 - (1 - O_h)^2) \cdot d_h} - \frac{q_h}{(1 - O_h)^2 \cdot d_h} \right) 
+ \delta \left( \frac{cs \cdot q'_o}{d_o + O_h \cdot d_h} - \frac{q_h}{(1 - O_h) \cdot d_h} \right) \right) \]

\[ O_h \left( \delta^2 (1 - cs) \frac{q'_o}{d_o + (1 - (1 - O_h)^2) \cdot d_h} + \delta \left( \frac{cs \cdot q'_o}{d_o + O_h \cdot d_h} - \frac{q_h}{(1 - O_h) \cdot d_h} \right) \right) \]

(170)
as compared to foragers that did not observe.

In addition the foragers that notice this change the round after it has happened will gain approximately

\[ \delta^2 (1 - cs) \frac{q'_o}{d'_o} + \delta \left( \frac{q'_o}{d_o + (1 - (1 - O_h)^2) \cdot d_h} - \frac{q_h}{(1 - O_h)^2 \cdot d_h} \right) \]

(171)
as compared to foragers that did not observe.

Thus an approximation of the expected total future payoff a forager gains from knowing the quality of the other patch when at a source patch of quality \( q_h \), can be computed given the functions \( d_{source} \) and \( O_{source} \). Similarly an approximation of the expected total future payoff a forager gains from knowing the quality of the other patch when at a sink patch of quality \( q_h \) can be computed given the functions \( d_{sink} \) and \( O_{source} \). A computer program that does this for each case can be found in appendix B.5.
Approximating The ES $d_{\text{source}}, O_{\text{sink}}$ and $O_{\text{source}}$

Now that we have a way of approximating the function $D(q_h, q_o)$, we can use this function to determine when it is worth switching from one patch to another. The cost of switching from a $q_o$ quality patch to a $q_h$ quality patch, when there will be $d_h = d_{\text{sink}}(q_h, q_o)$ of the foragers at the $h$ patch and $d_o = d_{\text{source}}(q_o, q_h)$ of the foragers at the $o$ patch is approximately $q_o/d_o - c_s \cdot q_h/d_h$. The benefit from switching in this case is approximately $D(q_h, q_o)$. Thus if $q_o/d_o - c_s \cdot q_h/d_h > D(q_h, q_o)$ then $d_{\text{source}}(q_h, q_o)$ is too low to satisfy the conditions of definition 5.4, and if $q_o/d_o - c_s \cdot q_h/d_h < D(q_h, q_o)$ then $d_{\text{source}}(q_h, q_o)$ is too high to satisfy the conditions of 5.4. Thus $d_{\text{source}}(q_h, q_o)$ (recall that $d_{\text{sink}}(q_o, q_h) = 1 - d_{\text{source}}(q_h, q_o)$) can be computed as the value that satisfies $q_o/d_o - c_s \cdot q_h/d_h \approx D(q_h, q_o)$.

Similarly since we can compute an approximation of the total expected future reward resulting from knowledge of the quality of the other patch, and we know that the cost of observing from a patch of quality $q_h$ with $d_h = d_{\text{source}}(q_h, q_o)$ or $d_h = d_{\text{sink}}(q_h, q_o)$ of the population at it is, as compared to not observing at that same patch, approximately $(1 - c_o)q_h/d_h$. If what is gained from observation is less than the cost of observation, then $O_{\text{source}}(q_h, q_o)$ (or $O_{\text{sink}}(q_h, q_o)$ as appropriate) is too high to satisfy the conditions of 5.7 (or the conditions of 5.9 as appropriate). On the other hand if what is gained from observation is greater than the cost of observation, then $O_{\text{source}}(q_h, q_o)$ (or $O_{\text{sink}}(q_h, q_o)$ as appropriate) is too low to satisfy the conditions of 5.7 (or the conditions of 5.9 as appropriate). Thus $O_{\text{source}}(q_h, q_o)$ and $O_{\text{sink}}(q_h, q_o)$ can be computed as the values which make it so that our approximation of what is gained from observations is equal to the our approximation of the cost of observation.
Figure 6: The equilibrium proportion of foragers at the source patch as a function of the quality of the sink patch values for source patches of value 1, 5, 10, 20, 50, and 100.

If we knew $d_{source}$ we would have a method of computing approximations of $O_{sink}$ and $O_{source}$. If we knew $O_{sink}$ and $O_{source}$ we would have a method of computing approximations of $d_{source}$. Since neither $d_{source}$ nor $O_{sink}$ and $O_{source}$ are known initially we use the following iterative method to compute approximations of these functions. We initially start with a guess of all these functions $O_{source0}$ and $O_{sink0}$, and $d_{source0}$. We then treat $O_{source0}$ and $O_{sink0}$ as the true $O_{source}$ and $O_{sink}$ functions, and use this to compute a new guess of $d_{source}$ called $d_{source1}$. Then we treat $d_{source1}$ as the true $d_{source}$ and use this to compute new guesses for $O_{source}$ and $O_{sink}$ called $O_{source1}$.
and \( O_{sink1} \). This process then proceeds iteratively until the guess functions have converged to some stable value. A computer program which implements this iterative process can be found in appendix B.5. We then use the functions that this iterative process converges on as our approximations of the function \( d_{source}, O_{source}, \) and \( O_{sink} \) of an ES policy.

![Contour plot of the ES probability of observation at the source patch as a function of source and sink patch quality.](image)

Figure 7: Contour plot of the ES probability of observation at the source patch as a function of source and sink patch quality.

This approximation of the equilibrium proportion of foragers at a source patch as a function of the sink patch quality, for select source patch qualities, is shown in figure 6. Contour plots of the approximate ES observation rates at the source and sink patches, as a function of source and sink patch quality are shown in figures 7 and
8. For these figures the parameters of the model are $P_c = 0.1$, $P_d = 0.02$, $\lambda = 0.1$, $c_o = 0.9$ and $c_s = 0.5$.

![Figure 8: Contour plot of the ES probability of observation at the sink patch as a function of source and sink patch quality.](image)

To get a sense of what figures 6, 8, and 7 are about consider the following. If a forager finds itself in a situation where the quality of its current patch is $q_h = 4$ and it knows that the other patch has quality $q_o = 8$ then according to figure 6 if the proportion of the population of foragers currently at the $h$ patch is greater than 0.36 this means that there are too many foragers at the $h$ patch and so foragers at the $h$ patch will have incentive to switch to the $o$ patch until about 0.36 of the population remains at the $h$ patch and 0.64 of the population is at the $o$ patch. Note that if
the population were simply dividing themselves up so that the payoffs per round per forager were equal there would be 0.33 of the population at the $h$ patch and 0.67 of the population at the $o$ patch. This 0.03 difference in the distributions is the result of foragers taking into account the costs of switching, and also the relative advantages of being at a patch with more or less foragers at it.

Suppose that this forager ends up switching to the $o$ patch along with other informed foragers from the $h$ patch and bringing about an active equilibrium distribution. Thus our forager is at a sink patch with quality $q_o = 8$ and sharing the resources there with $d_o = 0.64$ of the rest of the population. How frequently or with what probability should this forager observe the other patch, to see if it has improved and is now worth switching to? According to figure 8, on each round the forager should observe with probability $O_{\text{sink}}(8, 4) \approx 0.14$. If this forager had stayed behind at the quality 4 patch, then according to figure 7 they should be observing with probability $O_{\text{source}}(4, 8) \approx 0.05$.

5.2 Discussion

When the state is in equilibrium, a forager cannot improve its expected payoff by switching. These equilibrium situations are not permanent since the patch qualities are determined by a stochastic dynamic process, and so every now and then the system is perturbed from its equilibrium state by a change in patch quality. The approximate and incomplete policy we have found can be interpreted as the ES amount of effort that foragers should invest in stabilizing the state, so that the cost of stabilizing effort is compensated for by finding and exploiting disequilibrium opportunities.

Although a direct comparison between problem 3.1 and problem 5.1 is difficult, the
comparison of the following carefully contrived case serves to illustrate and interesting difference between the two problems. In problem 3.1, there was a single forager trying to optimize its reward intake in the face of patch quality dynamics identical to those faced by the population of foragers in this most recent problem, 5.1, where each individual in the population is trying to optimize their individual reward intake. In problem 3.1 a forager at a patch of quality 13, that believed the other patch was of quality 11, adopted an approximately optimal policy of checking in on the 11 quality patch after every 7 rounds of foraging at the 13 quality patch, that is $t^*_c(13, 11) \approx 7$.

Roughly equating this to the terms of problem 5.1, a lone forager in this situation will check in on the quality of the other patch one out of every eight rounds, that is with probability 0.125, and pay a cost of 2 reward units to do so. In problem 5.1 a forager at a sink patch of quality 8, where the source patch is of quality 12 and where there is $d_{sink}(8, 12) \approx 0.40$ at this sink patch, will observe each round with probability $O_{sink}(8, 12) \approx 0$, and pay a cost of 1.99 reward units to do so. Thus although the cost of observation is the same in both of these contrived situations in the different problems, the expected gains from observation are so low in the many forager problem that there is only incentive to observe with vanishing probability, where as in the lone forager problem there is incentive to “observe”, by which we mean switch to the other patch to check its value, every few rounds.

When a lone forager “observes,” the information it receives benefits it alone, and this benefit persists until a future patch change or the foraging session ending renders this information irrelevant. This is not so in the many forager case. In the many forager case when a forager observes, and the information it receives is beneficial, then this observing forager will act on this information by switching to the other patch.
This action then reveals to the patch mates of the forager that just observed, that the other patch has changed quality and is worth switching to. The patch mates of the observer receive this information without having to pay the cost of observation, and thus benefit from the observation made by their former patch mate. When the observing forager’s former patch mates act on the information revealed by the observing foragers’ actions, they bring the system into a new equilibrium distribution, rendering the benefit of the observers information irrelevant within a few rounds of when the observer initially acquired the information. Thus we should expect there to be dramatically less observation per forager in a many forager situation than in a lone forager situation.

A striking similar phenomenon has been studied in an economic context, by Grossman and Stiglitz “On the impossibility of informationally efficient markets” [11]. In this model agents can pay a cost to acquire information about the true vale of a risky asset, and use this information to inform their trading of this risky asset with others. In this economic model it is found that for certain parameter values it is beneficial for some traders to acquire this information, but not all, since the value of information lies in the advantage that it gives an informed agent over an uninformed agent. As in the problems of this thesis, the actions of the informed individuals in this model reveal information to the uninformed individuals. This creates something of a paradox where acting on information that was costly to acquire, reveals, at least in part the information to those that have not payed the cost to acquire the information initially, and thus diminishes the potential reward from having acquired the information in the first place.
Chapter 6

Concluding Remarks

In the previous chapters we worked through a specific set of model problems, yet the
dilemma explored by these particular problems, of whether to invest in information
about potential new opportunities or to take advantage of the opportunities at hand
is ubiquitous and fundamental.

6.1 The General Problem

Consider any spatially structured environment, with some temporal stochasticity,
such that the resource density throughout the environment changes in some unpredict-
dable way. Suppose that a population is distributed throughout this environment,
and that the fitness rewards to a forager are proportional to the level of environmental
resources in its location and inversely proportional to the density of fellow foragers
in that same location. We will say that a population distribution is in equilibrium,
if no member of the population would want to migrate to a new location in the
environment, given that they had perfect knowledge of the environmental resource
distribution and the population distribution. Members of this population allocate their effort amongst three activities, competing with their neighbors to exploit the local resources, migrating within the environment to new locals, and assessing the resource density at other locations within the environment. Assuming that assessing resource density returns more information per cost than migration, then when the population distribution is in equilibrium, effort will be allocated to assessment and exploitation, but not to migration. As a result this stable state will persist, until the temporal stochasticity of the resource density renders the current distribution of the population unstable. When the state is perturbed from its equilibrium, opportunities exist for foragers to temporarily gain an advantage over their peers by migrating to areas of the environment where the resource to forager ratio is better than at their current location, enough so that the costs of migration are offset. This advantage is temporary, because after some period of time the population will reach a stable state where the differences in resources per individual between any two areas is not significant enough to offset the costs of migration. How long this period of disequilibrium lasts is entirely dependent on how much effort the population as a whole allocates to assessing the quality of the environment, and how the information revealed by the actions of informed foragers spreads through the population. If lots of effort is allocated to assessing environment quality, then the population will learn of the changes in environment quality quickly, migrate to take advantage of quality discrepancies, and a stable state will be reached quickly. If very little effort is allocated to assessing environment quality, then the population will learn of the changes in the environment slowly, and so it will take a long time for those foragers migrating to take advantage of these quality discrepancies to stabilize the state. Thus in a population
prone to frequent observation and migration it would pay to observe and switch less frequently, and in a population that never observed or migrated it may pay to observe and migrate with relatively high frequency. The ES strategy is one where the level of observation and migration is such that when the whole population adopts this level of observation and migration no individual can do better by increasing or decreasing effort in observation or migration.

6.2 Many eyes make light work

In the many forager situation of problem 5.1 there seems to be less individual information gathering effort than in a lone forager situation like probe 3.1. We should expect the ES policy, in the many forager problems, to have lower individual observation effort and migration effort as the density of foragers increases, up to some limit. Consider that the addition of a single forager to a patch decreases the relative value of observation, since the chance of disequilibrium opportunities persisting for any length of time decreases with the number of foragers around to take advantage of these opportunities. Consider that in the lone forager case a disequilibrium opportunity persists until either the forager capitalizes on it, or it disappears as a result of further environmental change. As more foragers are introduced, if the frequency (or probability as appropriate) of observation is held fixed, then the probability of any particular forager being the one that capitalizes on an opportunity decreases, thus decreasing the value of observation. Thus we see a corresponding decrease in each individual’s level of observation effort.
6.3 What does an ES strategy optimize?

It is clear that competition necessarily leads to sub optimal resource use. Consider that in problem 5.1 if all of the foragers adopt a policy, call it $π^{fair}$, such that the foragers simply divide themselves evenly between the two patches and never observe and never switch, then the expected lifetime reward of a forager using an ES policy, $π^*$, in a population of foragers using $π^*$ would be the same as a foragers using $π^{fair}$ in a population using $π^{fair}$, less all of the costs of observing and switching paid by the foragers using $π^*$.

When looking for ES strategies we typically consider the perspective of an individual forager trying to optimize its fitness, given that the population was employing a given strategy. We then find the strategy that happens to be the best response to itself. It is interesting to take a step back from this individual based perspective and consider instead the goals of the genome of a population. To do this consider a population composed entirely of the grandchildren of matriarch $A$. This matriarch’s fitness is proportional to the average fitness of the population as a whole. If this matriarch were able to dictate which foraging strategy the population should use, what would she choose? For her the best case scenario is any strategy where the population doesn’t waste any effort on switching or observing. This however is not what the population actually does. The reason for this is that the population is not guaranteed to consist entirely of matriarch $A$’s grandchildren. If matriarch $A$’s grandchildren all adopt this cooperative strategy, they leave themselves vulnerable to being exploited by the less cooperative grandchildren of rival matriarch $B$.

Given this threat of matriarch $B$’s uncooperative grandchildren, how should matriarch $A$ change her grandchildren’s strategy to ward off this threat. Consider that
the situations an uncooperative forager can take advantage of are precisely those where the population is not in an equilibrium distribution. To defend against non-cooperators, matriarch A’s grandchildren must come up with a scheme for keeping the population distribution at equilibrium as often as possible, while at the same time not wasting too much effort on observing and switching. The (approximate) ES strategies that we have found are precisely the strategies that keep the population distribution in an equilibrium state frequently enough that there is nothing to be gained from switching or observing more than the ES strategy. Note that if a rare mutant forager were to never observe in a population playing the ES strategy, they would have the same expected fitness, as those employing the ES strategy. However, as soon as there is a significant proportion of the population observing and switching less than the ES levels, this population becomes vulnerable to invasion by rare mutants with higher levels of observation and switching.

Given that there is a cooperative strategy interesting questions arise about how inclusive fitness might affect this model. A particularly exciting extension of this work would be to incorporate an explicit death-birth (or birth-death) process into the model, and then consider the effects of various types of offspring dispersal between the two patches, and the relatedness structure that emerges.

6.4 Approximating Evolutionarily Stable Policies Using Value Differencing

To compute the (approximately) optimal and ES policies for problems 3.1, 3.2, 5.1, we compute the (approximate) difference in expected total future reward when taking
one action versus another. This is in contrast to the approach used in problems 4.1 and 4.2, of computing the actual expected total future reward from taking each possible action when is a particular state, using a value iteration algorithm. The structure of problems 3.2 and 5.1 is particularly well suited to this technique for the following reasons.

Consider that in order to choose optimally between two options in the present, a decision maker must consider the ramifications of this decision on into the future indefinitely. To choose approximately optimally, instead of considering the ramifications of this decision up until the end of time, a decision maker can consider only all of the ramifications of this present choice up to some finite time horizon. If the decision maker discounts future rewards then the importance of how this present decision affects the distant future becomes less important, which justifies this approximation technique and allows the loss from choosing an action based on this approximation versus choosing the optimal action to be bounded.

Suppose though that a forager does not discount future rewards. In a changing world, to which the decision maker responds, it is possible that a decision maker will find themselves in the same future situation, or at least in one of several equally valuable future situations, regardless of the particular decision made in the present situation. If the probability of a decision in the present being irrelevant in the future increases the further into the future a forager considers, this also justifies only considering the ramifications of the decision up to some finite time horizon, and allows the loss from choosing an action based on this approximation versus choosing the optimal action to be bounded.

Recall from problem 5.1 that when computing the expected total future payoff
difference from being at a sink patch of quality $q_h$ or a source patch of quality $q_o$, that we were able to exactly compute the expected total payoff difference prior to and during the first patch quality change or the session ending. Then conditioning on the type of the change, i.e. was it the session ending, or a patch quality changing, and if it was a patch quality changing, which of six possible types of patch quality change was it, and making assumptions about the policy employed by the population, and the population distribution dynamics that this policy produced and assuming that the contributions to this difference in the case where another patch quality change follows shortly behind the first patch quality change is small, we were able to approximate the contribution to this difference after the first change.

Recall that if a patch quality change occurred, it was of one of the six following types.

case 1. The $h$ patch improves to quality $q_h'$ so that $d_{source}(q_o, q_h') < d_o$, and so some foragers will be switching to the $h$ patch.

case 2. The $o$ patch improves to quality $q_o'$ so that $d_{source}(q_h, q_o') < d_h$, and so some foragers will be switching to the $o$ patch.

case 3. The $h$ patch worsens to quality $q_h'$ so that $d_{sink}(q_o, q_h') \leq d_o \leq d_{source}(q_o, q_h')$, and so even though there was a patch quality change no foragers will be switching from one patch to the other.

case 4. The $o$ patch improves to quality $q_o'$ so that $d_{sink}(q_h, q_o') \leq d_h \leq d_{source}(q_h, q_o')$, and so even though there was a patch quality change no foragers will be switching from one patch to the other.
case 5. The $h$ patch worsens to quality $q'_h$ so that $d_{source}(q'_h, q_o) < d_h$, and so some foragers will be switching to the $o$ patch.

case 6. The $o$ patch worsens to quality $q'_o$ so that $d_{source}(q'_o, q_h) < d_o$, and so some foragers will be switching to the $h$ patch.

Recall that in cases 1, 2, 5, and 6, the policy used by the population was assumed to be such that the response of the population brought about new active equilibrium distribution shortly after the patch quality change occurred, and thus we were able to exactly compute the expected total payoff difference from being at a $q_h$ quality sink patch versus a $q_o$ quality source patch, after the first patch quality change, assuming that there was not a second patch quality change while the population was responding to the first patch quality change.

Now in cases 3 and 4, no segment of the population switched patches in response to the patch quality change, and so in order to compute the payoff difference in these cases requires conditioning on the type of the second change to occur. Again at the second change there will be same six possible types of patch quality change, and for cases 1, 2, 5, and 6 an approximation of the resultant difference can be readily computed, but in cases 3 and 4 further conditioning on the third type of change to occur is required. In this problem it is clear that a decision made in the present will continue to affect a forager’s payoff into the future until, either the foraging session ends, or there is a patch quality change of type 1, 2, 5, or 6.

It is clear that both a high rate of discounting future rewards, and a high probability that at some future time a decision made in the present will cease to be relevant, serve to improve the probability that a choice of action made based on only considering the reward up to some finite time horizon will result in expected reward that is
close to the expected reward when the optimal choice of action is made. However it is unclear to me at this point how both discounting and the probability of irrelevance should be properly combined when computing approximations of an optimal policy, and bounding the loss due to using these approximation instead of the optimal choice of action.

Multi-agent extensions of Markov decision processes are in general difficult to deal with. However in the particular multi-agent problems investigated in this thesis, the extra structure imposed on the problems by considering the necessary qualities of ES policies in the infinite agent case, greatly simplified the problems. This is illustrated quite clearly by problems 3.1 and 3.2. In the lone forager case (problem 3.1) when using a value differencing technique, it was necessary to assume in many different cases that the difference in value between certain belief states was small. In contrast in the many forager case (problem 3.2) no such assumptions were necessary since the population’s response to patch quality changes, when using the ES policy, was such that decisions made in the past had no bearing on the future reward of a forager.

Thus there appears to be potential for an interesting synthesis between evolutionary game theory and Markov decision processes, where the nature of an ES policy is such that, as in problem 3.2 the relevance of a decision made in the present, ceases to have any effect shortly after that decision is made. Thus in cases where it may not be feasible to use a standard value iteration and replicator dynamics approach, as was used in problem 4.2, value differencing can be used to compute approximation of ES policy by considering the ramifications of a decision up to some point in the future where the expected difference in reward resulting from the present is small. Given the recent surge of interest in animal information use [7], and given that evolution is
an inherently game theoretic process, such a synthesis could have broad applications to the study of behaviour, ecology and evolution.
Bibliography


Appendix A

Supporting Calculations

A.1 When observation is preferable to innovation: almost always

In the introduction it is claimed that “learning through observation is in almost all cases better than learning through innovation.” The first qualification we apply to this claim is that we consider only the case where the sole motivation for a forager to observe or innovate is to learn about a new patch to exploit, (or in the case of observation, learn that a patch that the forager already knew about has changed to a new quality so that it is now worth exploiting). This assumption implies that the collateral benefits of improved estimates of the environmental state and its dynamics resulting from the information gained through observation or innovation, are insignificant compared to the direct benefits of potentially learning about a high quality patch that can then be exploited. Thus when deciding between observation and innovation the only factors to be considered are the distributions of the quality of a patch that will be observed as compared to the distribution of the quality of a patch that will be discovered through innovation. We then assume further that with regards to a forager’s goal of maximizing its average lifetime per round payoff, the higher moments of these patch quality distributions are unimportant, and that a forager will always do better on average, by choosing the learning option whose quality distribution has the highest mean.

Definition A.1. For a specific round of a given simulation we define the following.

$q_i$ denotes the quality of patch $i$, where $i$ indexes all of the patches.

$q_j$ denotes the quality of the patch that forager $j$ exploits on this round, where $j$ indexes all of the foragers that exploited this round.

$a_j$ denotes the set of patches that forager $j$ is aware of.

$a = \bigcup_j a_j$ denotes the set of patches that at least one forager is aware of.
\( \bar{a} \) denotes the set of patches that no exploiting forager is aware of.

In a slight stretch of standard notation we let \( E_j[E_a[q]] \) be denoted by \( E_a[q] \).

**Definition A.2.** The strategies of a population are said to be behaving pathologically in a given time step if, \( E_j[q] < E_{i \in a}[q] \), that is if on average the quality of a patch that a forager is choosing to exploit is less than the average quality of the patches that it is currently aware of and could be exploiting.

**Proposition A.3.** The average quality of a patch discovered through innovation being greater than the average quality of an observable patch that is, \( E_i[q] > E_j[q] \), implies that either

1. That the average quality of the patches that the foragers are aware of, weighted by how many foragers are aware of each patch, is less than the average quality of the patches that the foragers are unaware of, \( E_a[q] < E_{\bar{a}}[q] \).

2. That the strategies of the population are behaving pathologically, \( E_j[q] < E_a[q] \).

**Proof.** \( E_a[q] \geq E_{\bar{a}}[q] \) implies that \( E_a[q] \geq E_i[q] \).

Thus if \( E_j[q] \geq E_a[q] \) then \( E_i[q] \leq E_j[q] \). □

While many strategy profiles may occasionally produce states where the population is engaging in pathological behaviour, given that a strategy’s frequency within the population is proportional to the average lifetime per round reward of a forager, strategies that engage in pathological behaviour frequently will likely not ever compose a large part of the population in a tournament simulation. Thus we can expect the situations where condition 2 of proposition A.3 holds to be rare.

Since the stochastic process of patch quality changes does not differentiate between patches that foragers are aware of and patches that foragers are not aware of, the changing of patch qualities serves diminish any systematic difference between \( E_a[q] \) and \( E_{\bar{a}}[q] \) that may exist as a result of the foragers’ strategies. Thus for high rates of \( P_c \) it can be expected that \( E_a[q] \approx E_{\bar{a}}[q] \).

In low \( P_c \) environments it is possible that the learning strategies of the foragers will create systematic differences between \( E_a[q] \) and \( E_{\bar{a}}[q] \), the question is will \( E_a[q] < E_{\bar{a}}[q] \) or will \( E_a[q] \geq E_{\bar{a}}[q] \)?

If foragers primarily learn about new patches through innovation then of course \( E_a[q] \approx E_{\bar{a}}[q] \).

If foragers learn about new patches primarily through observation there are two important consequences for the difference between \( E_a[q] \) and \( E_{\bar{a}}[q] \).

1. When there is more observation and less innovation the size of the set \( a \) will be smaller than in the case where there is some innovation, since the only way the set \( a \) can grow is through innovation and errors in observation. If the size of \( a \)
APPENDIX A. SUPPORTING CALCULATIONS

is less than half the number of patches in the environment, then as the size of $a$
decreases, the variance, in time, of $E_a[q] - \bar{E}_a[q]$ increases. This is because when
most of the population is aware of the same patch $i$ then a large change in this
patch $i$’s quality will drastically affect $E_a[q]$. When $a$ is small it is necessarily
the case that many foragers are aware of the same patches.

2. When there is more observation and less innovation, and the population is not
engaged in pathological behaviour, since the expected quality of the patches
observed is equal to the average quality of the patches exploited $E_j[q]$, and
$E_j[q] \geq E_a[q]$, it is clear that $E_a[q] - \bar{E}_a[q]$ when considered as a time series will
have a non-negative trend.

If the variance in time of $E_a[q] - \bar{E}_a[q]$ due to the non-negative trend is large
compared to undirected variance caused by patch quality changes, then we can expect
condition 1 of proposition A.3 to hold rarely. If on the other hand the undirected variance in time of $E_a[q] - \bar{E}_a[q]$ due to patch quality changes is on the same scale as
the variance in time due to the non-negative trend, then the occasions where condition
1 of proposition A.3 can potentially be frequent.

The results of the tournament [20] showed that very low levels of innovation or very
low levels of observation errors resulting in the discovery of a new patch, were sufficient
to maintain $a$ at a size where the undirected variance of $E_a[q] - \bar{E}_a[q]$ due to patch
quality changes was small compared to the variance in $E_a[q] - \bar{E}_a[q]$ due to the non-
negative trend. Thus while it is not the case that innovation is strictly dominated by
observation, the situations where the expected quality of a patch discovered through
innovation is higher than the expected quality of an observed patch are rare, and
further these situations only come about when foragers that innovate are very rare
or absent.
A.2 Conditioning on the time and type of the first and second changes in problem 3.1

Consider foragers that leave a patch of quality \( q_h \) to check in on a patch that was of quality \( q_o \) the last time this other patch’s quality was known with certainty. Suppose that one of these foragers checks in on the \( q_o \) quality patch after \( t_c \) rounds of exploiting at the \( q_h \) quality patch, and the other forager checks in after every \( t_c + 1 \) rounds of exploiting at a \( q_h \) quality patch. In section 3.1.3, the difference in expected lifetime payoff between these two foragers prior to the first change has already been computed. In what follows an approximation of the difference in expected lifetime payoff during and after this first change occurs is computed by conditioning on the timing and type of the first and second changes. The notation used is the same as is used throughout the first section of chapter 3.

Before calculating this contribution to \( D \) the following definitions are required.

**Definition A.4.** Let \( r_l \) (l for learn) be the number of rounds that will pass after the \( q_o \) quality patch changes, before a forager using a policy of checking in on the \( q_o \) quality patch after every \( t_c \) rounds will discover that the \( q_o \) quality patch has changed, provided that the foraging session doesn’t end or the \( q_h \) quality patch does not change prior to when the forager is scheduled to check in on the \( q_o \) quality patch.

\[
r_l = t_c - ((r_1 - 1) \mod (t_c + 1))
\]

\( r_l \) is a random variable, with its distribution depending on the distribution of \( r_1 \) and the forager’s strategy. \( r_1 \) is distributed on \([0, 1, 2, \ldots, t_c] \) like a truncated, inverted geometric random variable, so that

\[
\frac{\delta^{t_c-k}(1-\delta)}{1-\delta^{t_c+1}} \quad (173)
\]

\( r_l \) can be thought of as the number of rounds that the forager will remain ignorant of the change in quality at the \( q_o \) quality patch. If \( r_l = 0 \) this means that the forager was checking in on the \( q_o \) quality patch right when the \( q_o \) quality patch changed. If \( r_l = t_c \) this means that the \( q_o \) quality patch changed right after the forager checked in on it, and so another \( t_c \) rounds will pass before the forager checks in on it again. We let \( r'_l \) be the same quantity as \( r_l \) except for a forager with a policy of checking in on the \( q_o \) quality patch after every \( t_c - 1 \) rounds at the \( q_h \) quality patch, giving

\[
r'_l = t_c - 1 - ((r_1 - 1) \mod (t_c)).
\]

For convenience we let \( m \) denote the minimum of \( r_l \) and \( r'_l \), which is the earliest time a forager can have discovered the change in the \( q_o \) quality patch for this particular
APPENDIX A. SUPPORTING CALCULATIONS

value of $r_1$, and let $M$ denote the maximum of $r_l$ and $r'_l$, which is the latest time a forager can have discovered the change in the $q_o$ quality patch. Note that on average $r'_l < r_l$ but that in some cases $r_l = r'_l$ and in some cases $r_l < r'_l$.

**Definition A.5.** Let $r_2$ be the time after the first change occurs, given that the first change is the $q_o$ quality patch changing, that a second distinguishable change could be noticed by a forager using one of the two policies.

For example the $q_o$ quality patch changing quality several times, prior to a forager using either policy learning that the $q_o$ quality patch has changed, is not significant, whereas the foraging session ending is always significant. $\gamma = (1 - P_d)(1 - P_c/2)$ denotes the probability that the $q_h$ quality patch doesn’t change and the session doesn’t end on a given round. The distribution of $r_2$ can be described as follows.

In the case where the first change is the $q_o$ quality patch changing to $q' \leq q_h$, then: For $r_2 < M$, a significant change, that is the foraging session ending or the $q_h$ quality patch changing, occurs $r_2$ rounds after the first change with probability $\gamma^{r_2-1}(1 - \gamma)$. For $r_2 = M$ a significant change occurs $M$ rounds after the first change with probability $\gamma^{M-1}(1 - \delta)(1 - P_c' M - m)$. Here $P_c'' = \frac{P_c}{2 - P_c}$ denotes the per round probability of the $q_o$ quality patch changing again, given that the $q_h$ quality patch has not changed.

In the case where the first change is the $q_o$ quality patch changing to $q' > q_h$, then: For $r_2 \leq m$, a significant change, that is the foraging session ending or the $q_h$ quality patch changing, occurs $r_2$ rounds after the first change with probability $\gamma^{r_2-1}(1 - \gamma)$. For $m < r_2 \leq M$ a significant change occurs $r_2$ rounds after the first change with probability $\gamma^m \delta^{r_2-m-1}(1 - \delta)$.

With these definitions in hand we are now ready to compute an approximation of the contribution to $D$ occurring on the round of the first change and after.

Case A. The first change to occur is the foraging session ending. This occurs with probability

$$\frac{P_d}{P_d + P_c - P_d P_c} \quad (175)$$

The contribution to $D$ is of course zero.

Case B. The first change to occur is the $q_h$ quality patch changing. This occurs with probability

$$\frac{P_c (1 - P_d)}{2(P_d + P_c - P_d P_c)} \quad (176)$$
APPENDIX A. SUPPORTING CALCULATIONS

Case B.1. Using either policy a forager will be at the $q_o$ quality patch when the $q_h$ quality patch changes. This happens when $r_1 \bmod (t_c(t_c + 1)) = 0$ which occurs with probability

$$\frac{\delta^{(t_c+1)(t_c)-1}(1 - \delta)}{1 - \delta^{(t_c+1)(t_c)}}$$

(177)

The contribution to $D$ is of course zero.

Case B.2. Using either policy a forager will be at the $q_h$ quality patch when it changes. This happens when $r_1 \bmod (t_c) \neq 0$ and $r_1 \bmod (t_c + 1) \neq 0$. Letting $\alpha = \delta^{t_c+1}$ and $\beta = \delta^{t_c}$ this occurs with probability

$$1 - \left(\frac{1 - \delta}{1 - \delta^{(t_c+1)(t_c)}}\right) \left(\frac{(1 - \delta)}{\delta^{(t_c+1)(t_c)-1}} - \frac{(\alpha - \alpha^{t_c})}{\delta(1 - \alpha)(1 - \alpha^{t_c})} - \frac{(\beta - \beta^{t_c+1})}{\delta(1 - \beta)(1 - \beta^{t_c+1})}\right)$$

(178)

The contribution to $D$ is

$$E[V(q', q_o, r_1 \bmod (t_c + 1)) - V(q', q_o, r_1 \bmod (t_c))| q_h \text{ changed to } q' = q_o] = 0$$

(179)

We assume that this is negligible.

Case B.3. Using a policy of checking less frequently, that is after every $t_c$ rounds, results in being at the $q_o$ quality patch when the $q_h$ quality patch changes, where as using a policy of checking more frequently, that is after every $t_c - 1$ rounds, results in being at the $q_h$ quality patch when it changes. This happens when $r_1 \bmod (t_c + 1) = 0$ and $r_1 \bmod (t_c) \neq 0$ and occurs with probability

$$\frac{(1 - \delta)(\alpha - \alpha^{t_c})}{\delta(1 - \alpha)(1 - \alpha^{t_c})}$$

(180)

Case B.3.1. The new value of the (formerly) $q_h$ quality patch is $q' = q_o$. This occurs with probability $\lambda(1 - \lambda)^{q_o-1}$. The contribution to $D$ is

$$E[V(q_o, q_h, 1) - V(q_o, q_o, r_1 \bmod (t_c))| q_h \text{ changed to } q' = q_o] = 0$$

(181)

Case B.3.2. The new value of the (formerly) $q_h$ quality patch satisfies $q' >
This occurs with probability \((1 - \lambda)^{q_o}\). The contribution to \(D\) is
\[
E[(q_o - q') + (V(q_o, q_o, 1) - V(q', q_o, r_1 \text{ mod } (t_c)))| \text{given } q_h \text{ has changed to } q' > q_o]
\]
\[
\approx E[q_o - q'| q' > q_o] = -1/\lambda
\]
(182)

We assume that the expectation of \(V(q_o, q_h, 1) - V(q', q_o, r_1 \text{ mod } (t_c))\) given the conditions of this case is negligible.

Case B.3.3. The new value of the (formerly) \(q_h\) quality patch satisfies \(q' < q_o\). This occurs with probability \(1 - (1 - \lambda)^{q_o-1}\).

Case B.3.3.1. There is no change immediately following the first change.

This occurs with probability \(\delta\). The contribution to \(D\) is
\[
E[(q_o - q') + (q' - q_o) + (V(q', q_o, 1) - V(q_o, q', 1))| q' < q_o] \approx 0
\]
(183)

We assume that the expectation of \(V(q', q_o, 1) - V(q_o, q', 1)\) given the conditions of this case is negligible.

Case B.3.3.2. There is a change immediately following the first change, and it is the foraging session ending. This occurs with probability \(P_d\). The contribution to \(D\) is
\[
E[(q_o - q')| q' < q_o]
\]
(184)

Case B.3.3.3. There is a change immediately following the first change and it is the quality of the \(q_o\) quality patch changing to quality \(q''\). This occurs with probability \((1 - P_d)(P_e/2)\). The contribution to \(D\) is
\[
E[(q_o - q') + (q' - q'') + (V(q', q_o, 1) - V(q'', q', 1))| \text{given } q_h \text{ changed to } q' < q_o \text{ and } q_o \text{ changed to } q'']
\]
\[
\approx E[q_o - q'']
\]
(185)

We assume that the expectation of \(V(q', q_o, 1) - V(q'', q', 1)\) given the conditions of this case is negligible.

Case B.3.3.4. There is a change immediately following the first change and it is the quality of the (formerly) \(q_h\) quality patch changing
again. This occurs with probability \( (1 - P_d)(P_c/2) \). The contribution to \( D \) is

\[
E[[q_o - q'] + (q'' - q_o) + (V(q'', q_o, 1) - V(q_o, q', 1))] \\
given q' changed to q'' and q' < q_o \\
\approx E[q'' - q'|q' < q_o] \quad (186)
\]

We assume that the expectation of \( V(q'', q_o, 1) - V(q_o, q', 1) \) given the conditions of this case is negligible.

Case B.4. Using a policy of checking less frequently, that is after every \( t_c \) rounds, results in being at the \( q_h \) quality patch when it changes, where as using a policy of checking more frequently, that is after every \( t_c - 1 \) rounds results in being at the \( q_o \) quality patch when the \( q_h \) quality patch changes. This happens when \( r_1 \mod (t_c + 1) = 0 \) and \( r_1 \mod (t_c) \neq 0 \) and occurs with probability

\[
\frac{(1 - \delta)(\beta - \beta^{t_c+1})}{\delta(1 - \beta)(1 - \beta^{t_c+1})} \quad (187)
\]

Case B.4.1. The new value of the (formerly) \( q_h \) quality patch is \( q' = q_o \). This occurs with probability \( \lambda(1 - \lambda)^{q_o-1} \). The contribution to \( D \) is

\[
E[V(q_o, q_o, r_1 \mod (t_c + 1) - V(q_o, q_h, 1))] \mid q_h \text{ changed to } q' = q_o \approx 0 \quad (188)
\]

We assume that this contribution is negligible.

Case B.4.2. The new value of the (formerly) \( q_h \) quality patch satisfies \( q' > q_o \). This occurs with probability \( (1 - \lambda)^{q_o} \). The contribution to \( D \) is

\[
E[(q' - q_o) + (V(q', q_o, r_1 \mod (t_c + 1) - V(q_o, q_h, 1))] \\
given q_h \text{ changed to } q' > q_o \\
\approx E[q' - q_o|q' > q_o] = 1/\lambda \quad (189)
\]

We assume that the expectation of \( V(q', q_o, r_1 \mod (t_c+1)) - V(q_o, q_h, 1) \) given the conditions of this case is negligible.

Case B.4.3. The new value of the (formerly) \( q_h \) quality patch satisfies \( q' < q_o \). This occurs with probability \( 1 - (1 - \lambda)^{q_o-1} \).

Case B.4.3.1. There is no change immediately following the first change.
This occurs with probability $\delta$. The contribution to $D$ is

$$ E[(q' - q_o) + (q_o - q') + (V(q_o, q', 1) - V(q', q_o, 1))|q' < q_o] \approx 0 $$

(190)

We assume that the expectation of $V(q_o, q', 1) - V(q', q_o, 1)$ given the conditions of this case is negligible.

Case B.4.3.2. There is a change immediately following the first change, and it is the foraging session ending. This occurs with probability $P_d$. The contribution to $D$ is

$$ E[(q' - q_o)|q' < q_o] $$

(191)

Case B.4.3.3. There is a change immediately following the first change and it is the quality of the $q_o$ quality patch changing to quality $q''$. This occurs with probability $(1 - P_d)(P_c/2)$. The contribution to $D$ is

$$ E[(q' - q_o) + (q'' - q') + (V(q'', q', 1) - V(q', q_o, 1))|$$

given $q_h$ changed to $q'$ and $q_o$ changed to $q''$]

$$ \approx E[q'' - q_o] $$

(192)

We assume that the expectation of $V(q'', q', 1) - V(q', q_o, 1)$ given the conditions of this case is negligible.

Case B.4.3.4. There is a change immediately following the first change and it is the quality of the (formerly) $q_h$ quality patch changing again. This occurs with probability $(1 - P_d)(P_c/2)$. The contribution to $D$ is

$$ E[(q' - q_o) + (q_o - q'') + (V(q_o, q', 1) - V(q'', q_o, 1))|$$

given $q'$ has changed to $q''$ and $q' < q_o$]

$$ \approx E[q' - q''|q' < q_o] $$

(193)

We assume that the expectation of $V(q_o, q', 1) - V(q'', q_o, 1)$ given the conditions of this case is negligible.

Case C. The first change to occur is the $q_o$ quality patch changing. This occurs with probability

$$ \frac{P_c(1 - P_d)}{2(P_d + P_c - P_dP_c)} $$

(194)
Case C.1. The $q_o$ quality patch changes to $q' \leq q_h$. In this case it is not nearly so critical to discover that the $q_o$ quality patch has changed. This happens with probability $1 - (1 - \lambda)^{q_h}$.

Case C.1.1. The $q_h$ quality patch changes or the foraging session ends before a forager using either policy will learn of the $q_o$ quality patch changing. This happens when $r_2 < m$, which occurs with probability

$$1 - \gamma^{m-1} \text{ when } m > 0$$
$$0 \text{ when } m = 0$$

(195)

Case C.1.1.1. The change is the foraging session ending. This happens with probability

$$P_d$$

$$P_d + P_c/2 - P_dP_c/2$$

(196)

The contribution to $D$ in this case is 0.

Case C.1.1.2. The change is the $q_h$ quality patch changing to $q''$. This happens with probability

$$\frac{P_c(1 - P_d)}{2(P_d + P_c/2 - P_dP_c/2)}$$

(197)

The contribution to $D$ in this case is

$$E[V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q'', q_o, r_2 + r_1 \mod (t_c)) | q_o \text{ changed to } q' \leq q_h \text{ and } q_h \text{ changed to } q'']$$

$$\approx 0$$

(198)

We assume this contribution to $D$ to be negligible.

Case C.1.2. The $q_h$ quality patch changes or the foraging session ends at the earliest time a forager will discover that the $q_o$ quality patch has changed. This happens when $r_2 = m$, which occurs with probability

$$\gamma^{m-1}(1 - \gamma) \text{ when } m > 0$$
$$0 \text{ when } m = 0$$

(199)

Case C.1.2.1. The change is the foraging session ending. This happens
with probability

$$\frac{P_d}{P_d + P_c/2 - P_dP_c/2}$$

(200)

The contribution to $D$ in this case is 0.

Case C.1.2.2. The change is the $q_h$ quality patch changing. This happens with probability

$$\frac{P_c(1 - P_d)}{2(P_d + P_c/2 - P_dP_c/2)}$$

(201)

The contribution to $D$ in this case is

$$E[((q' - q'') + V(q', q_o, 1) - V(q'', q_o, r_2 + r_1 \text{ mod } (t_c))) \Pr(r_l < r'_l)$$

$$+ ((q'' - q') + V(q'', q_o, r_2 + r_1 \text{ mod } (t_c + 1)) - V(q', q_h, 1)) \Pr(r'_l < r_l)]$$

given $q_o$ changed to $q' \le q_h$ and $q_o$ changed to $q''$

$$\approx E[((q' - q'') \Pr(r_l < r'_l) + (q'' - q') \Pr(r'_l < r_l)|q' \le q_h]$$

(202)

We assume that the expectations of

$$(V(q', q_h, 1) - V(q'', q_o, r_2 + r_1 \text{ mod } (t_c))) \Pr(r_l < r'_l)$$

and

$$(V(q'', q_o, r_2 + r_1 \text{ mod } (t_c + 1)) - V(q', q_h, 1)) \Pr(r'_l < r_l)$$

are negligible, given the conditions of this case.

Case C.1.3. There is a patch quality change or the foraging session ends after the earliest time that a forager will discover that the $q_o$ quality patch has changed, but before the latest time that a forager will discover that the $q_o$ quality patch has changed. This happens when $m < r_2 < M$ which occurs with probability

$$\gamma^m(1 - \gamma^{M-m-1}) \text{ when } M - m > 0$$

$$0 \text{ when } M - m = 0$$

(203)

Case C.1.3.1. The change is the foraging session ending. This happens with probability

$$\frac{P_d}{P_d + P_c/2 - P_dP_c/2}$$

(204)
In this case the contribution to $D$ is

$$E[(q' - q_h) \Pr(r_l < r'_l) + (q_h - q') \Pr(r'_l < r_l)|q' \leq q_h] \quad (205)$$

Case C.1.3.2. The change is the $q_h$ quality patch changing to quality $q''$. This occurs with probability

$$\frac{P_c(1 - P_d)}{2(P_d + P_c/2 - P_dP_c/2)} \quad (206)$$

In this case the contribution to $D$ is

$$E[((q' - q_h) + V(q'', q', r_2 - m) - V(q''', q_o, r_l + r_1 \text{ mod } (t_c))) \Pr(r_l < r'_l)$$

$$+ ((q_h - q') + V(q'', q_o, r'_l + r_1 \text{ mod } (t_c + 1)) - V(q''', q', r_2 - m)) \Pr(r'_l < r_l)]$$

given $q_o$ changed to $q'' \leq q_h$ and $q_h$ changed to $q''$]

$$\approx E[(q' - q_h) \Pr(r_l < r'_l) + (q_h - q') \Pr(r'_l < r_l)|q' \leq q_h] \quad (207)$$

We assume that the expectations of

$$V(q'', q', r_2 - m) - V(q''', q_o, r_l + r_1 \text{ mod } (t_c))) \Pr(r_l < r'_l)$$

and

$$V(q'', q_o, r'_l + r_1 \text{ mod } (t_c + 1)) - V(q''', q', r_2 - m)) \Pr(r_l > r'_l)$$

are negligible given the conditions of this case. Throughout we will be taking the expectation of the quantities $r_2 - m$ and $M - m$. In this case and many others to follow the expectation of these quantities should be taken conditional on $r_l < r'_l$ or $r'_l < r_l$. For clarity we suppress the explicit conditioning, since it is clear from context when these conditional expectations should be taken. For example, in this case we have written $V(q'', q', r_2 - m)$ instead of $E[V(q'', q', r_2 - m)|r_l < r'_l]$ and $E[V(q'', q', r_2 - m)|r'_l < r_l]$.

Case C.1.4. There is a patch quality change or the foraging session ends at the latest time that a forager will discover that the $q_o$ quality patch has changed. This patch quality change could potentially be the formerly $q_o$ quality patch changing for a second time, in the period between $m$ and $M$ rounds after the first change. This case happens when $m < r_2$.
and \( r_2 = M \) which occurs with probability
\[
\gamma^{M-1}(1 - (\delta)(\zeta)) \quad \text{when } M > m \\
0 \quad \text{when } M = m 
\]  
(208)

Where \( \zeta = (1 - P''_c)^{M-m-1} \) denotes the probability that the \( q_o \) quality patch has not changed again more than \( m \) rounds after the first change, but fewer than \( M \) rounds after the first change, and \( P''_c = P_c / (2 - P_c) \) is the probability of the \( q_o \) quality patch changing each round given that the \( q_h \) quality patch did not change on those rounds.

Case C.1.4.1. The change is the foraging session ending. This occurs with probability
\[
\frac{P_d}{1 - (\delta)(\zeta)} 
\]  
(209)

In this case the contribution to \( D \) is
\[
E[(q' - q_h) \Pr(r_l < r'_l) + (q_h - q') \Pr(r'_l < r_l)] | q' > q_h \]  
(210)

Case C.1.4.2. The change is the \( q_h \) quality patch changing to quality \( q'' \), while the formerly \( q_o \) quality patch has not changed again in the time between rounds \( m \) and \( M \) after the first change. This occurs with probability
\[
\frac{(P_c/2)(1 - P_d)(\zeta)}{1 - (\delta)(\zeta)} 
\]  
(211)

In this case the contribution to \( D \) is
\[
E[((q' - q_h) + (q'' - q') + (V(q'', q', M - m) - V(q', q_h, 1))) \Pr(r_l < r'_l) + ((q_h - q') + (q' - q'') + (V(q', q_h, 1) - V(q'', q', M - m))) \Pr(r'_l < r_l)] \\
given q_o changed to \( q' \) and \( q_h \) changed to \( q'' \) \\
\approx E[(q'' - q_h) \Pr(r_l < r'_l) + (q_h - q'') \Pr(r'_l < r_l)] 
\]  
(212)

We assume that
\[
(V(q'', q', M - m) - V(q', q_h, 1)) \Pr(r_l < r'_l) 
\]
and
\[(V(q', q_h, 1) - V(q'', q', M - m)) \Pr(r'_l < r_l)\]
are negligible given the conditions of this case.

Case C.1.4.3. The change is the formerly \(q_o\) quality patch changing again to quality \(q''\), any time between \(m + 1\) and \(M\) rounds, inclusive, after the first change. This occurs with probability
\[
\left(\frac{P_c}{2} + (1 - \zeta) - \frac{P_c}{2}(1 - \zeta)\right)(1 - P_d)
\]
In this case the contribution to \(D\) is
\[
E[((q' - q_h) + (q_h - q'')) \Pr(r_l < r'_l) + V(q_h, q', M - m) - V(q'', q_h, 1)) \Pr(r'_l < r_l) + (q_h - q') + (q'' - q_h) + V(q'', q_h, 1) - V(q_h, q', M - m)) \Pr(r'_l < r_l) | \text{given } q_o \text{ changed to } q' > q_h \text{ and changed again to } q'']
\approx E[(q' - q'') \Pr(r_l < r'_l) + (q'' - q') \Pr(r_l > r'_l) | \text{given } q_o \text{ changed to } q' \leq q_h \text{ and changed again to } q'']
\]
We assume that the expectations of
\[
(V(q_h, q', M - m) - V(q'', q_h, 1)) \Pr(r_l < r'_l)
\]
and
\[
(V(q'', q_h, 1) - V(q_h, q', M - m)) \Pr(r'_l < r_l)
\]
are negligible under the conditions of this case.

Case C.1.4.4. The change is the \(q_h\) quality patch changing to \(q''\) \(M\) rounds after the first change has occurred, and the \(q_o\) quality patch changing to \(q'''\) after \(m\) rounds after the first change has occurred, but before \(M\) rounds after the first change. This occurs with probability
\[
\left(\frac{P_c}{2}\right)(\zeta)(1 - P_d)
\]
\[
\frac{1 - (\delta)(\zeta)}{1 - (\delta)(\zeta)}
\]
(215)
The contribution to $D$ in this case is

$$E[\Pr(r_l < r'_l)((q' - q_h) + (q'' - q'''')$$

$$+ V(q'', q', M - m) - V(q''', q_h, 1))$$

$$+ \Pr(r'_l < r_l)((q_h - q') + (q''' - q'')$$

$$+ V(q''', q_h, 1) - V(q'', q', M - m)))]$$

given $q_o$ changed to $q' \leq q_h$ and then to $q'''$ and $q_h$ changed to $q''$

$$\approx E[((q' - q_h) \Pr(r_l < r'_l) + (q_h - q') \Pr(r'_l < r_l)|q' \leq q_h]$$  \hspace{1cm} (216)

We assume that

$$(V(q'', q', M - m) - V(q''', q_h, 1)) \Pr(r_l < r'_l)$$

and

$$(V(q''', q_h, 1) - V(q'', q', M - m)) \Pr(r'_l < r_l)$$

are negligible given the conditions of this case.

Case C.1.5. No significant change occurs until after a forager discovers that the $q_o$ patch has changed, regardless of which policy is used. A significant change is either the session ending or the $q_h$ quality patch changing at any time, or the $q_o$ quality patch changing yet again after the earliest time a forager can have discovered the first change. This happens with probability

$$\gamma_m \delta^{M-m}$$  \hspace{1cm} (217)

The contribution to $D$ in this case is

$$E[((q' - q_h) + (q_h - q')$$

$$+ (V(q_h, q', M - m) - V(q', q_h, 1)) \Pr(r_l < r'_l)$$

$$+ (V(q', q_h, 1) - V(q_h, q', M - m)) \Pr(r'_l < r_l)]$$

given $q_o$ changed to $q' \leq q_h] \approx 0$$  \hspace{1cm} (218)

We assume that the expectations of

$$V(q_h, q', M - m) - V(q', q_h, 1)$$

and

$$V(q', q_h, 1) - V(q_h, q', M - m)$$

are negligible under the conditions of this case. Note that if $q_o <$
\( q' < q_h \) it is possible that a forager that learns of the \( q_o \) quality patch changing to \( q' \) early, will check in on the \( q' \) quality patch again (or possibly several more times) before a forager that learns of the \( q_o \) quality patch change later checks in on the formerly \( q_o \) quality patch for the first time. In this case the above expression will be incorrect, with there potentially being more \((q' - q_h) \Pr(r_l < r'_l)\) and more \((q_h - q') \Pr(r'_l < r_l)\) terms and the \( V(q_h, q', M - m) \) terms having smaller values than \( M - m \) in their third argument. We assume however that the probability of these situations and the margin of error they introduce is small enough to be ignored.

Case C.2. The \( q_o \) quality patch changes to \( q' > q_h \). This occurs with probability \((1 - \lambda)^{q_h}\). This is the case that motivates checking in on the lower quality patch.

Case C.2.1. The \( q_h \) quality patch changes or the foraging session ends before a forager using either policy will learn of the \( q_o \) quality patch changing. This happens when \( r_2 < m \). This happens with probability

\[
1 - \gamma^{m-1} \quad \text{when } m > 0 \\
0 \quad \text{when } m = 0
\]  

(219)

Case C.2.1.1. The change is the foraging session ending. This happens with probability

\[
\frac{P_d}{P_d + P_c/2 - P_dP_c/2}
\]  

(220)

The contribution to \( D \) in this case is 0.

Case C.2.1.2. The change is the \( q_h \) quality patch changing to \( q'' \). This happens with probability

\[
\frac{P_c(1 - P_d)}{2(P_d + P_c/2 - P_dP_c/2)}
\]  

(221)

The contribution to \( D \) in this case is

\[
E[V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q'', q_o, r_2 + r_1 \mod (t_c))]|/\text{given } q_o \text{ changed to } q' > q_h \text{ and } q_h \text{ changed to } q''] \\
\approx 0
\]  

(222)

We assume this contribution to \( D \) to be negligible.

Case C.2.2. The \( q_h \) quality patch changes or the foraging session ends at
the earliest time a forager will discover that the \( q_o \) quality patch has changed. This happens with probability

\[
\gamma^{m-1}(1 - \gamma) \text{ when } m > 0 \\
0 \text{ when } m = 0
\]  

(223)

Case C.2.2.1. The change is the foraging session ending. This happens with probability

\[
\frac{P_d}{P_d + P_c/2 - P_dP_c/2}
\]  

(224)

The contribution to \( D \) in this case is 0.

Case C.2.2.2. The change is the \( q_h \) quality patch changing to \( q'' \). This happens with probability

\[
\frac{P_c(1 - P_d)}{2(P_d + P_c/2 - P_dP_c/2)}
\]  

(225)

The contribution to \( D \) in this case is

\[
E[[(q' - q'') + V(q', q_h, 1) - V(q'', q_o, r_2 + r_1 \mod (t_c))] \Pr(r_l < r'_l) \\
+ ((q'' - q') + V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q', q_h, 1)) \Pr(r'_l < r_l)] \\
given \( q_o \) changed to \( q'' \) and \( q_h \) changed to \( q'' \)
\approx E[(q' - q'') \Pr(r_l < r'_l) + (q'' - q') \Pr(r'_l < r_l)]
\]  

(226)

We assume that the expectations of

\[
(V(q', q_h, 1) - V(q'', q_o, r_2 + r_1 \mod (t_c))) \Pr(r_l < r'_l)
\]

and

\[
V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q', q_h, 1)) \Pr(r'_l < r_l)
\]

are negligible given the conditions of this case.

Case C.2.3. There is a patch quality change or the foraging session ends after the earliest time that a forager will discover that the \( q_o \) quality patch has changed, but before the latest time that a forager will discover that the \( q_o \) quality patch has changed. This occurs when
\[ m < r_2 < M \] which happens with probability
\[
\gamma^m (1 - \delta^{M-m-1}) \text{ when } M - m > 0
\]
\[ 0 \text{ when } M - m = 0 \] (227)

Case C.2.3.1. The change is the foraging session ending. This occurs with probability
\[
\frac{P_d}{P_d + P_c - P_d P_c} \] (228)

In this case the contribution to \( D \) is
\[
E[(q' - q_h)(r_2 - m) \Pr(r_l < r'_l) + (q_h - q')(r_2 - m) \Pr(r'_l < r_l)|q' > q_h] \] (229)

Case C.2.3.2. The change is the \( q_h \) quality patch changing to quality \( q'' \). This occurs with probability
\[
\frac{P_c(1 - P_d)}{2(P_d + P_c - P_d P_c)} \] (230)

In this case the contribution to \( D \) is
\[
E[((q' - q_h)(r_2 - m) + (q' - q'')) + V(q', q_h, r_2 - m + 1) - V(q'', q_o, r_2 + r_1 \mod (t_c)) \Pr(r_l < r'_l) + ((q_h - q')(r_2 - m) + (q'' - q')) + V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q', q_h, r_2 - m + 1)) \Pr(r'_l < r_l)|\]
given \( q_o \) changed to \( q' > q_h \) and \( q_h \) changed to \( q'' \]
\[
\approx E[((q' - q_h)(r_2 - m) + (q' - q'')) \Pr(r_l < r'_l) + ((q_h - q')(r_2 - m) + (q'' - q')) \Pr(r'_l < r_l)|\]
given \( q_o \) changed to \( q' > q_h \) and \( q_h \) changed to \( q'' \] (231)

We assume that
\[
(V(q', q_h, r_2 - m + 1) - V(q'', q_o, r_2 + r_1 \mod (t_c))) \Pr(r_l < r'_l)
\]
and
\[
(V(q'', q_o, r_2 + r_1 \mod (t_c + 1)) - V(q', q_h, r_2 - m + 1)) \Pr(r'_l < r_l)\]
are negligible given the conditions of this case.

Case C.2.3.3. The change is the formerly $q_o$ quality patch changing again to quality $q''$. This occurs with probability

$$\frac{P_c(1 - P_d)}{2(P_d + P_c - P_dP_c)}$$

(232)

In this case the contribution to $D$ is

$$E[((q' - q_h)(r_2 - m) + (q'' - q_h)$$
$$+ V(q'', q_o, r_2 - m + 1) - V(q_h, q_o, r_2 + r_1 \mod (t_c))) \Pr(r_i < r_i')$$
$$+ ((q_h - q')(r_2 - m) + (q_h - q'')$$
$$+ V(q_h, q_o, r_2 + r_1 \mod (t_c + 1)) - V(q'', q_h, r_2 - m + 1)) \Pr(r_i' < r_i)$$

given $q_o$ changed to $q' > q_h$ and changed again to $q''$]

$$\approx E[((q' - q_h)(r_2 - m) + (q'' - q_h)) \Pr(r_i < r_i')$$
$$+ ((q_h - q')(r_2 - m) + (q_h - q'')) \Pr(r_i' < r_i)$$

given $q_o$ changed to $q' > q_h$ and changed again to $q''$]

(233)

We assume that

$$(V(q'', q_h, r_2 - m + 1) - V(q_h, q_o, r_2 + r_1 \mod (t_c))) \Pr(r_i < r_i')$$

and

$$(V(q_h, q_o, r_2 + r_1 \mod (t_c + 1)) - V(q'', q_h, r_2 - m + 1)) \Pr(r_i > r_i')$$

are negligible under the conditions of this case.

Case C.2.4. There is a patch quality change or the foraging session ends, after the earliest time a forager will discover the change, and at the latest time that a forager will discover that the $q_o$ quality patch has changed. This occurs when $m < r_2$ and $r_2 = M$ which happens with probability

$$\gamma^m \delta^{M-m-1}(1 - \delta) \text{ when } M > m$$
$$0 \text{ when } M = m$$

(234)

Case C.2.4.1. The change is the foraging session ending. This occurs
APPENDIX A. SUPPORTING CALCULATIONS

with probability

\[
\frac{P_d}{P_d + P_c - P_dP_c}
\]  

(235)

In this case the contribution to \( D \) is

\[
E[(q' - q_h)(M - m) \Pr(r_l < r'_l) + (q_h - q')(M - m) \Pr(r'_l < r_l)|q' > q_h]
\]

(236)

Case C.2.4.2. The change is the \( q_h \) quality patch changing to quality \( q'' \). This occurs with probability

\[
\frac{P_c(1 - P_d)}{2(P_d + P_c - P_dP_c)}
\]

(237)

In this case the contribution to \( D \) is

\[
E[((q' - q_h)(M - m) + V(q', q_h, M - m + 1) - V(q', q_h, 1)) \Pr(r_l < r'_l) + ((q_h - q')(M - m) + V(q', q_h, 1) - V(q', q_h, M - m + 1)) \Pr(r'_l < r_l)]
\]

given \( q_o \) changed to \( q' > q_h \) and \( q_h \) changed to \( q'' \)

\[
\approx E[((q' - q_h)(M - m) \Pr(r_l < r'_l) + (q_h - q')(M - m) \Pr(r'_l < r_l)|q' > q_h]
\]

(238)

We assume that

\[
(V(q', q_h, M - m + 1) - V(q', q_h, 1)) \Pr(r_l < r'_l)
\]

and

\[
(V(q', q_h, 1) - V(q', q_h, M - m + 1)) \Pr(r'_l < r_l)
\]

are negligible given the conditions of this case.

Case C.2.4.3. The change is the formerly \( q_o \) quality patch changing again to quality \( q'' \). This occurs with probability

\[
\frac{P_c(1 - P_d)}{2(P_d + P_c - P_dP_c)}
\]

(239)
In this case the contribution to $D$ is

$$E[((q' - q_h)(M - m) + V(q'', q_h, M - m + 1) - V(q'', q_h, 1)) \Pr(r_l < r'_l)$$
$$+ ((q_h - q')(M - m) + V(q', q_h, 1) - V(q', q_h, M - m + 1)) \Pr(r'_l < r_l)]$$

given $q_o$ changed to $q' > q_h$ and changed again to $q''$]

$$\approx E[((q' - q_h)(M - m) \Pr(r_l < r'_l)$$
$$+ (q_h - q')(M - m) \Pr(r'_l < r_l)|q' > q_h]$$  \hspace{1cm} (240)

We assume that the expectations of

$$(V(q''', q_h, M - m + 1) - V(q'', q_h, 1)) \Pr(r_l < r'_l)$$

and

$$(V(q'', q_h, 1) - V(q'', q_h, M - m + 1)) \Pr(r'_l < r_l)$$

are negligible under the conditions of this case.

Case C.2.5. No significant change occurs until after a forager discovers that the $q_o$ patch has changed, regardless of which policy is used. A significant change is either the session ending or the $q_h$ quality patch changing at any time, or the $q_o$ quality patch changing yet again after the earliest time a forager can have discovered the first change. This happens with probability

$$\gamma^m \delta^{M-m}$$  \hspace{1cm} (241)

The contribution to $D$ in this case is

$$E[((q' - q_h)(M - m) + V(q', q_h, M - m + 1) - V(q', q_h, 1)) \Pr(r_l < r'_l)$$
$$+ ((q_h - q')(M - m) + V(q', q_h, 1) - V(q', q_h, M - m + 1)) \Pr(r'_l < r_l)|q' > q_h]$$

$$\approx E[(q' - q_h)(M - m) \Pr(r_l < r'_l) + (q_h - q')(M - m) \Pr(r'_l < r_l)|q' > q_h]$$  \hspace{1cm} (242)

We assume that the expectations of

$$(V(q', q_h, M - m + 1) - V(q', q_h, 1)) \Pr(r_l < r'_l)$$

and

$$V(q', q_h, 1) - V(q', q_h, M - m + 1)$$

are negligible under the conditions of this case.
A.3 The Policy and Belief State Dynamics of Problem 4.2

What follows is a complete description of the belief state transition process experienced by a forager using a candidate policy, in problem 4.2 with switching and observing quotas $S^t$ and $O^t$ with $t \in [1, 2, \ldots, T]$, in a population using that same candidate policy. Note that since this is a candidate policy $S^t = O^t = 0$ if $t \mod 2 = 0$.

Case F. When $\hat{q}_o < 0$ then

$$\pi(q_h, \hat{q}_o, d_h, t, d^t) = \begin{cases} 
-1 & \text{with probability } S^{t+1}/d_h \\
0 & \text{with probability } O^{t+1}/d_h \\
1 & \text{with probability } 1 - S^{t+1}/d_h - O^{t+1}/d_h
\end{cases} \quad (243)$$

If $t + 1$ is odd then:

If the forager switches its next belief state will be

$$b = (q_y, q_x, 1 - d^*(t + 1), t + 1, -1). \quad (244)$$

Here $q_y$ has probability $\lambda(1 - \lambda)^{q + \hat{q}_o}$ of having value $q$, where $q_y$ must be greater than $-\hat{q}_o$.

If the forager observes its next belief state will be

$$b = (q_x, q_y, d^*(t + 1), t + 1, 0). \quad (245)$$

Again $q_y$ has probability $\lambda(1 - \lambda)^{q + \hat{q}_o}$ of having value $q$, where $q_y$ must be greater than $-\hat{q}_o$.

If the forager exploits its next belief state will be

$$b = (q_x, \hat{q}_o, d^*(t + 1), t + 1, 1). \quad (246)$$

If $t + 1$ is even then the forager would only ever exploit:

If

$$q_y \leq \frac{q_h(1 - d_h)}{d_h \cdot g(t + 1)} \quad (247)$$

then the next belief state is

$$b = (q_x, 0, d^*(t), t + 1, 1). \quad (248)$$
If
\[
\frac{qh \cdot (1 - dh)}{dh \cdot g(t + 1)} < q_y < \frac{qh \cdot (1 - d'(t + 1))}{d'(t + 1) \cdot g(t + 1)}
\] (249)
then the next belief state is
\[
b = (qx, qy, \frac{qx}{qx + qy \cdot g(t + 1)}, t + 1, 1).
\] (250)

If
\[
\frac{qh \cdot (1 - d'(t + 1))}{d'(t + 1) \cdot g(t + 1)} \leq q_y
\] (251)
then the next belief state is
\[
b = (qx, \frac{qh \cdot (1 - d'(t + 1))}{d'(t + 1) \cdot g(t + 1)}, d'(t + 1), t + 1, 1) \] (252)

Case G. In the case where \( \hat{q}_o = 0 \)
\[
\pi(q_h, \hat{q}_o, d_h, t, a^t) = 1
\] (253)
In this case a forager will only ever exploit so the next belief state is
\[
b = (q_h, 0, d_h, t + 1, 1).
\] (254)

Case H. Let
\[
P_x = \min \left[ \max \left[ 0, \left( \frac{1}{Ot} \right) \left( d_h - \frac{qh}{q_h + q_o \cdot g(t + 1)} \right) \right], 1 \right]
\] (255)
In the case where \( \hat{q}_o > 0 \) and \( a^t = 0 \) then
\[
\pi(q_h, \hat{q}_o, d_h, t, a^t) = \begin{cases} 
-1 & \text{with probability } P_x \\
1 & \text{with probability } 1 - P_x
\end{cases}
\] (256)
If the forager switches patches the next belief state is
\[
b = (q_y, qx, 1 - d_h + Ot \cdot P_x, t + 1, -1) \] (257)
If the forager stays at its current patch the next belief state is
\[ b = (q_x, q_y, d_h - O^t \cdot P_x, t + 1, 1) \] (258)

Case I. Let
\[ P_y = \min \left[ \max \left[ 0, \left( \frac{1}{d_h} \right) \left( d_h + S^{t+1} - \frac{q_h}{q_h + q_o \cdot g(t + 1)} \right) \right], 1 \right] \] (259)

In the case where \( \hat{q}_o > 0 \) and \( a^t \neq 0 \) then
\[ \pi(q_h, \hat{q}_o, d_h, t, a^t) = \begin{cases} -1 \text{ with probability } P_y & \\ 1 \text{ with probability } 1 - P_y & \end{cases} \] (260)

If the forager switches back to the \( x \) patch the next belief state is
\[ b = (q_x, q_y, 1 - d_h - S^{t+1} + d_h \cdot P_y, t + 1, -1) \] (261)

If the forager stays at the \( y \) patch the next belief state is
\[ b = (q_y, q_x, d_h + S^{t+1} - d_h \cdot P_y, t + 1, 1) \] (262)

All of the foragers begin in belief state
\[ b = (q_x, -1, 1, 0, 1). \] (263)

The dynamics of the belief states of a rare mutant forager using a policy different from the rest of the population so that \( S^i \neq S^i_m \) or \( O^i \neq O^i_m \) for some \( i \in [1, 2, \ldots, T] \), will be identical to the dynamics described above, except in Case A with \( \hat{q}_o < 0 \) where the mutant forager will use \( S^i_m \) and \( O^i_m \) instead of \( S^i \) and \( O^i \) when determining its probability to switch, exploit or observe.
A.4 The Policy at a Sink Patch for problem 5.1

In section 5.1.2, the policy of a forager as determined by the functions \( d_{\text{source}}^{i}, d_{\text{sink}}^{i}, O_{\text{source}}^{i}, S_{\text{source}}^{i}, S_{\text{sink}}^{i}, O_{\text{sink}}^{i}, W_{\text{leap}} \) and \( W_{\text{look}} \), is described in detail for those cases when a forager is a source patch, and then a change in quality occurs and the population of foragers responds to this change. Those cases described concerning foragers also hold for foragers at a sink patch, with minor modification, and are described in detail here.

Case F. The forager’s patch was a sink patch the last time there was an active equilibrium and the forager’s patch has not changed quality since that time, that is \( q_{h} = q_{h}^{\text{e}} \) and \( d_{h} = d_{\text{sink}}(q_{h}^{\text{e}}, q_{o}^{\text{e}}) \). So far as the forager knows the other patch has not changed to anything worth switching to, or from. That is if the other patch has changed quality to \( q_{o}^{\prime} \), the following is satisfied, \( d_{\text{sink}}(q_{h}, q_{o}^{\prime}) \leq d_{h} \leq d_{\text{source}}(q_{h}, q_{o}^{\prime}) \), which places bounds on \( q_{o}^{\prime} \). In this case the forager will observe with probability \( O_{\text{sink}}(q_{h}^{\text{e}}, q_{o}^{\text{e}}) \). If the forager does not observe then it will exploit. The forager will never switch patches in this case.

Case G. The forager’s patch was a sink patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. For compactness we let \( O = O_{\text{sink}}(q_{h}^{\text{e}}, q_{o}^{\text{e}}) \), \( d = d_{\text{sink}}(q_{h}^{\text{e}}, q_{o}^{\text{e}}) \), \( O' = O_{\text{source}}(q_{h}, q_{o}^{\prime}) \) and \( d' = d_{\text{source}}(q_{h}, q_{o}^{\prime}) \). The forager knows that the other patch has just changed to something worth switching to, \( q_{o}^{\prime} \), because that forager was one of \( O \cdot d \) foragers at the patch that observed this round. In this case the forager will migrate from this patch with probability \( P_{m} \), such that if possible the population distribution will be in active equilibrium on the next round, that is

\[
d - d' = d \cdot O \cdot P_{m} \tag{264}
\]

and so solving for \( P_{m} \) we have

\[
P_{m} = \min \left[ 1, \frac{d - d'}{d \cdot O} \right] \tag{265}
\]

If the forager stays at this patch, which it will do with probability \( 1 - P_{m} \) then the forager will observe with probability \( P_{o} \) so that if possible

\[
d' \cdot O' = d' \cdot (1 - O) \cdot O + d' \cdot (1 - O) \cdot (1 - P_{m}) \cdot (P_{o}) \tag{266}
\]
and so solving for $P_o$ we have,

$$P_o = \max \left[ 0, \min \left[ 1, \frac{d' \cdot O' - d \cdot (1 - O) \cdot O}{d \cdot (1 - O) \cdot (1 - P_m)} \right] \right]$$  \hspace{1cm} (267)$$

Thus a forager in this case takes action $a = -1$ with probability $P_m$, takes action $a = 0$ with probability $(1 - P_m)P_o$ and takes action $a = 1$ with probability $(1 - P_m)(1 - P_o)$.

Case H. The forager’s patch was a sink patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. For compactness we let $O^i = O^i_{\text{sink}}(q^x_h, q^x_o)$, $S^i_{\text{sink}}(q^x_h, q^x_o)$, $d = d_{\text{sink}}(q^x_h, q^x_o)$, $d' = d_{\text{source}}(q_h, q'_o)$, and $O' = O_{\text{source}}(q_h, q'_o)$. Further we let

$$d^n = d \cdot \prod_{i=0}^{n} (1 - S^i - O^i)$$

denote the proportion of the total population remaining at the former sink patch, that have not observed the other patch during the $n$ rounds since the quality of the former source patch improved enough to cause migration, given that the population is not yet in an equilibrium distribution. The forager knows that the other patch has changed to something worth switching to on previous rounds because all of the foragers that observed on the previous rounds have switched to the other patch. The forager is one of the $O^n \cdot d^{n-1}$ foragers remaining at the patch that observed the quality of the other patch on this current round. In this case, the other foragers at this patch, though they do not know the precise quality of the other patch, are aware that the quality of the other patch has improved enough that there is incentive to switch there, and so $S^{n+1} \cdot d^n$ foragers will blindly switch to the other patch on next round. This informed forager takes this into account, and migrates to the other patch with probability $P_m$, so that if possible the population distribution is in an active equilibrium on the next round, that is

$$d^n \cdot (1 - S^{n+1}) - d' = d^{n-1} \cdot O^n \cdot P_m$$  \hspace{1cm} (268)$$

and so solving for $P_m$ we have,

$$P_m = \max \left[ 0, \min \left[ 1, \frac{d^n \cdot (1 - S^{n+1}) - d'}{d^{n-1} \cdot O^n} \right] \right]$$  \hspace{1cm} (269)$$

If some, but not all of these informed foragers remains behind, that is $0 < P_m < \ldots$
1, they will observe with probability $P_o$, so that if possible
\[ O' \cdot d' = O^{n+1} \cdot d^n + O^n \cdot d^{n-1} \cdot (1 - P_m) \cdot P_o \] (270)

Solving for $P_o$ we have that
\[ P_o = \max \left[ 0, \min \left( 1, \frac{O' \cdot d' - O^{n+1} \cdot d^n}{O^n \cdot d^{n-1} \cdot (1 - P_m)} \right) \right] \] (271)

The probability with which these informed foragers should observe in the case where $P_m = 0$, cannot be described using just the functions $d_{\text{source}}$, $d_{\text{sink}}$, $O_{\text{source}}^i$, $O_{\text{sink}}^i$, $S_{\text{source}}^i$, $S_{\text{sink}}^i$, $W_{\text{look}}$ and $W_{\text{leap}}$.

Case I. The forager’s patch was a sink patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. For compactness we let $O^i = O_{\text{sink}}^i(q_e, q_o)$, $d = d_{\text{sink}}(q_h, q_o)$, $O' = O_{\text{source}}(q_h, q_o)$ and $d' = d_{\text{source}}(q_h, q_o)$. Further we let
\[ d^n = d \cdot \prod_{i=0}^{n} (1 - S^i - O^i) \]

denote the proportion of the total population remaining at the former source patch, that has not observed the other patch, $n$ rounds after the quality of the former sink patch improved enough to cause migration, and given that the population is not yet in an equilibrium distribution. The forager knows that the other patch has changed to something worth switching to on previous rounds because all of the foragers that observed on the previous round have switched to the other patch after observing the quality of the other patch. The forager is not one of the $O^n \cdot d^{n-1}$ foragers remaining that has observed the quality of the other patch on this current round. In this case, the forager is aware that the quality of the other patch has improved enough that there is incentive to switch there, and so the forager will observe with probability $O^{n+1}$, switch to the other patch with probability $S^{n+1}$ and exploit with probability $1 - O^{n+1} - S^{n+1}$.

Case J. The forager’s patch was a sink patch the last time there was an active equilibrium and the forager’s patch quality has not changed since that time. A forager has observed that the other patch just worsened to $q'_o$, and so knows that $(1 - d) \cdot W_{\text{leap}}$ foragers will be switching to its patch, and that this switching will bring about an active equilibrium. For compactness we let $O = O_{\text{sink}}(q_h, q'_o)$, $d = d_{\text{sink}}(q_h, q'_o)$, $O' = O_{\text{sink}}(q_h, q'_o)$ and $d' = d_{\text{sink}}(q_h, q'_o)$. The forager will
observe with probability $P_o$ such that if possible

$$O' \cdot d' = d \cdot (1 - O) \cdot O + d \cdot O \cdot P_o$$  \hspace{1cm} (272)

Solving for $P_o$ we have that

$$P_o = \max \left[ 0, \min \left[ 1, \frac{O' \cdot d' - d \cdot (1 - O) \cdot O}{d \cdot O} \right] \right]$$  \hspace{1cm} (273)
Appendix B

Computer Programs

B.1 Program for computing an approximation of $t^*_c(q_h, q_o)$

The code that follows is documented well enough to be self explanatory. This first piece of code computes an approximation of $D$ using the conditioning described in the preceding appendix and in section 3.1.3. Using this approximation of $D$ a second program then computes an approximation of the optimal frequency of checking in on the other patch, $t^*_c$.

```matlab
function dif = incredif(t,qh,qo,Pd,Pc,lambda)
    %This function computes the expected payoff difference between a forager
    %checking in on the qo quality patch after every t rounds of foraging at a
    %qh quality patch versus a forager checking in after every t-1 rounds when
    %faced with a qh and a qo patch, with qo<qh. This is done by conditioning
    %on the time and type of the first few changes.
    delta = (1-Pd)*(1-Pc);
    gamma = (1-Pd)*(1-Pc/2);
    dif = 0;
    %The direct benefit of checking in on the low quality patch less frequently
    %is
    benefitlesschecking = (qh-qo)*(delta^t*(1-delta))...
```

184
\[ \frac{1}{(1-\delta^{t+1})*(1-\delta^t)}; \]

dif = dif + benefitlesschecking;

%r1 is the number of rounds until the first change occurs. It is
%geometrically distributed with parameter delta. The value of r1 mod
%t*(t-1) lets us know whether it is the forager that is checking every t
%rounds or a forager checking every t-1 rounds that will discover a change
%first, and how many rounds it will take them to discover it.

index = 1:t*(t+1);
for r1 = index

  %This is the probability of this particular value of r1 mod t*(t-1)
  %occurring.
  probr1 = (1-delta)*delta^(r1-1)/(1-delta^(t*(t+1)));

%rl1 is the number of rounds after the first change, if that first
%change is the qo quality patch changing, until the forager checks qo
%quality patch, given that they check in on the low quality patch every
%t rounds. It can be thought of as the number of rounds that the
%forager remains ignorant to the change occuring.
  rl1 = t - mod(r1-1,t+1);

%rl2 is the same quantity as rl1 except for a forager checking in after
%every t-1 rounds.
  rl2 = (t-1) - mod(r1-1,t);

%The earliest time after the first change that a forager will check in
%on the qo quality patch
  m = min(rl1,rl2);
  M = max(rl1,rl2);

if m == M
  %If the forager will check in on the (possibly formerly) qo quality
  %patch at the same time regardless of which policy is used, then
  %clearly there will be no further contribution to the difference
  %between the two policies.
  continue
end;
%From this point on we only consider the situation where M is strictly greater than m.

%Pcprime is the per round probability of the formerly qo quality patch changing given that the qh quality patch has not changed on those rounds.
Pcprime = Pc/(2-Pc);

%zeta is the probability of the qo quality patch not changing for a second time after m rounds after the first change has passed but before M rounds after the first change has passed, given that the session has not ended and the qh patch has not changed quality in that time.
zeta = (1-Pcprime)^(M-m-1);

%We need to consider the change type, is it the session ending or the qh or the qo patch quality changing. We don’t consider the case where the first change is the session ending since of course this does not lead to any further difference in the performance of the two policies.
for changetype1 = 1:2
    if changetype1 == 1 % qh is the patch that changes
        %This is the probability of the qh patch changing, and the session not ending given that a change has occurred
        probchangetype1 = Pc*(1-Pd)/2/(Pd + Pc - Pd*Pc);
    end
    %It is then convienent to consider when a forager using the either of the two policies will discover that the change has occured
    if rl1 == rl2
        %In this case since both foragers will check in on the qo patch at the same time we consider the further difference to be negligible.
    elseif rl1 == 0
        %This means that a forager that checks after every t rounds will be checking in on the qo patch when the qh patch changes quality, but not a forager that checks after every t-1 rounds.
    end
    %We then need to consider whether the new value of the qh patch, q', satisfies q'<qo, or q'=qo, or q'>qo. If q'=qo there will be no significant further contribution to the difference.
for qualitytype1 = 1:2
if qualitytype1 == 1
    %The probability that q' will have this type of
    %quality is
    probqualitytype1 = 1 - (1-lambda)^(qo-1);

    %In this case the qh patch changes to something bad
    %right when a forager checking after every t rounds
    %is checking in on the qo quality patch. The
    %expected new quality (enq) is

    %If this case is not possible, i.e. qo == 1 then enq cannot
    %be computed sensibly.
    if probqualitytype1 > 0
        tmp = 1:(qo-1);
        enq = (tmp * ((1-lambda).^(tmp-1))')... 
             /sum((1-lambda).^(tmp-1));
    else
        enq = 1;
    end;

end;

%We then need to consider whether there is a second
%change immediately following the first change.
for changetype2 = 1:4
if changetype2 == 1
    %There is no immediate second change.
    probchangetype2 = delta;
    dif = dif + 0;
elseif changetype2 == 2
    %The immediate second change is the session
    %ending.
    probchangetype2 = Pd;
    dif = dif ... 
        + probr1 * probchangetype1... 
        * probqualitytype1 * probchangetype2... 
        * (qo-enq);
elseif changetype2 == 3
    %The immediate second change is the qh patch
    %changing quality.
    probchangetype2 = (1-Pd)*Pc/2;
dif = dif + ...
    + probr1*probchangetype1...
    * probqualitytype1 * probchangetype2...
    * (1/lambda - enq);
elseif changetype2 == 4
    %The immediate second change is the qo patch
    %changing quality
    probchangetype2 = (1-Pd)*Pc/2;
    dif = dif + ...
        + probr1 * probchangetype1...
        * probqualitytype1 * probchangetype2...
        * (qo - 1/lambda);
end; %changetype2 if loop
end; %changetype2 for loop
elseif qualitytype1 == 2 %q'>qo
    probqualitytype1 = lambda^(qo);
    %In this case the qh patch changes to something
    %good. The expected new quality (enq) is
    enq = qo + 1/lambda;
    dif = dif ...
        + probr1 * probchangetype1 * probqualitytype1...
        * (qo - enq);
end;
end; %qualitytype1 for and if loops
elseif rl2 == 0
    %This means that a forager that checks after every t-1 rounds
    %will be checking in on the qo patch when the qh patch changes.
    %We then need to consider whether the new value of the qh
    %patch, q', satisfies q'<qo, or q'=qo, or q'>qo. If q'=qo there
    %will be no significant further contribution to the difference.
    for qualitytype1 = 1:2
        if qualitytype1 == 1 %q'<qo
            probqualitytype1 = 1 - lambda^(qo-1);
            %In this case the qh patch changes to something bad right
            %when the a forager checking after every t-1 rounds is
            %checking in on the qo quality patch. The expected new
            %quality (enq) is
if probqualitytype1 > 0
    tmp = 1:(qo-1);
    enq = (tmp * ((1-lambda).^(tmp-1)).') / sum((1-lambda).^(tmp-1));
else
    enq = 1;
end;

% We then need to consider whether there is a second change immediately following the first change.

for changetype2 = 1:4
    if changetype2 == 1
        % There is no second change immediately following the first change.
        % probchangetype2 = delta;
        dif = dif + 0;
    elseif changetype2 == 2
        % The immediate second change is the session ending.
        probchangetype2 = Pd;
        dif = dif + ...
            + probr1 * probchangetype1...
            * probqualitytype1 * probchangetype2...
            * (enq-qo);
    elseif changetype2 == 3
        % The immediate second change is qh patch changing quality.
        probchangetype2 = (1-Pd)*Pc/2;
        dif = dif + ...
            + probr1 * probchangetype1...
            * probqualitytype1 * probchangetype2...
            * (enq - 1/lambda);
    elseif changetype2 == 4
        % The immediate second change is qo patch changing quality.
        probchangetype2 = (1-Pd)*Pc/2;
dif = dif + ...
    + probr1 * probchangetype1...
    * probqualitytype1 * probchangetype2...
    * (1/lambda - qo);
end;
end; %changetype2 for and if loops
elseif qualitytype1 == 2 %q'>qo
    probqualitytype1 = lambda^(qo);
    %In this case the qh patch changes to something good. The
    %expected new quality (enq) is
    enq = qo + 1/lambda;
    dif = dif ...
        + probr1 * probchangetype1 * probqualitytype1...
        *(enq - qo);
end;
end; %qualitytype1 for and if loops
else %rl1 ~= 0 and rl2 ~= 0
    %in this case regardless of the policy used a forager will be
    %aware of the qh quality patch change right when it happens, we
    %assume that the differences after this change are negligible.
end; %rl1 versus rl2 if loop
elseif changetype1 == 2 %the qo quality patch changes to quality q'
    probchangetype1 = Pc*(1-Pd)/2/(Pd + Pc - Pd*Pc);
    %We next need to consider whether the qo patch gets better than qh
    %or worse.
    for qualitytype1 = 1:2
        if qualitytype1 == 1 %q'<=qh
            probqualitytype1 = 1-(1-lambda)^qh;
            %In this case the qo patch has changed to something not so
            %fantastic, and so it is not so great to learn about it
            %changing. The expected new quality (enq) of the formerly qo
            %quality patch in this case is
            tmp = 1:qh;
            enq = (tmp * ((1-lambda).^(tmp-1))')...
                /sum((1-lambda).^(tmp-1));
        %We next need to consider when the second change occurs
        for r2 = 1:(M+1)
if r2 < M  
  probr2 = gamma^(r2-1)*(1-gamma);  
elseif r2 == M  
  probr2 = gamma^(M-1) * (1-delta*zeta);  
elseif r2 == M+1  
  %we let r2 == M+1 represent all r2 values greater than  
  %M  
  probr2 = gamma^(M-1) * delta * zeta;  
end;

if r2 < m  
  %The second change occurs before a forager using either  
  %policy would discover that the qo quality patch has  
  %changed. That is r2 < m
  
  %If the foraging session ends then there is no further  
  %contribution to dif.  
  
  %If the qh quality patch changes then there is no  
  %significant contribution to dif, since a forager will  
  %be in the state (q'',qo,t1) compared to (q'',qo,t2).  
  dif = dif + 0;

elseif r2 == m  
  %The second change occurs right as a forager using the  
  %early policy discovers the change in qo
  
  %If this change is the session ending then there is no  
  %further contribution to dif, however if this change is  
  %the qh quality patch changing there is a contribution  
  %to dif. The qh quality changing is the second change  
  %with probability
  probchangetype2 = ...  
    Pc*(1-Pd)/2/(Pd+Pc/2-Pd*Pc/2);  

  %We then need to consider whether the forager checking  
  %more or less frequently will be first to discover the  
  %change for this particular value of r1.  
  if r11 < r12  
    %A forager checking after every t rounds will learn
%about the qo change before a forager checking
%after every t-1 rounds
dif = dif + probr1*probchangetype1...
   *probqualitytype1*probchangetype2...
   *(enq - 1/lambda);
elseif rl2 < rl1
    %A forager checking after every t-1 rounds will
    %learn about the qo change before a forager
    %checking after every t rounds.
    dif = dif + probr1*probchangetype1...
        *probqualitytype1*probchangetype2...
        *(1/lambda - enq);
end;
%Note that if rl1 = rl2 there is no difference

elseif and(m<r2, r2<M)
    %The second change occurs after a forager using the
    %early policy discovers the qo change, but before a
    %forager using the late policy discovers the change.
    %Since qo has changed to something bad an early forager
    %will switch back right away, and then shortly after a
    %change will occur.
    if rl1 < rl2
        dif = dif + probr1 * probchangetype1...
            *probqualitytype1 * probchangetype2 * (enq - qh);
    elseif rl2 < rl1
        dif = dif + probr1 * probchangetype1...
            *probqualitytype1 * probchangetype2 * (qh - enq);
    end;
elseif r2 == M
    %The second change occurs right as a forager using the
    %late policy discovers that the qo quality patch has
    %changed.
    %We then need to consider the type of this change
for changetype2 = 1:4
  if changetype2 == 1
    %This change is the session ending
    probchangetype2 = Pd/(1-delta*zeta);
    
    if r1 < r2
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (enq - qh);
    elseif r2 < r1
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (qh - enq);
    end;
  elseif changetype2 == 2
    %This change is the qh patch changing
    probchangetype2 = Pc/2*(1-Pd)*zeta/(1-delta*zeta);
    
    if r1 < r2
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (1/lambda - qh);
    elseif r2 < r1
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (qh - 1/lambda);
    end;
  elseif changetype2 == 3
    %This change is the qo patch changing again
    probchangetype2 = (Pc/2*(1-Pd)*zeta ...
      +(1-zeta)*delta)... /
    /(1-delta*zeta);
    
    if r1 < r2
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (enq - 1/lambda);
    elseif r2 < r1
      dif = dif + probr1 * probchangetype1...
      * probqualitytype1 * probr2...
      * probchangetype2 * (enq - 1/lambda);
\[ \text{dif} = \text{dif} + \text{probr1} \times \text{probchangetype1} \times \text{probqualitytype1} \times \text{probr2} \times \text{probchangetype2} \times \left( \frac{1}{\lambda} - \text{enq} \right); \]

\text{end;}

\text{elseif changetype2 == 4}
\%
\text{This change is the qo patch changing again and the qh patch changing}
\text{probchangetype2} = \frac{P_c/2*(1-P_d)*(1-zeta)}{(1-delta*zeta)};
\%
\text{if r1 < r2}
\text{dif} = \text{dif} + \text{probr1} \times \text{probchangetype1} \times \text{probqualitytype1} \times \text{probr2} \times \text{probchangetype2} \times (\text{enq} - \text{qh});
\text{elseif r2 < r1}
\text{dif} = \text{dif} + \text{probr1} \times \text{probchangetype1} \times \text{probqualitytype1} \times \text{probr2} \times \text{probchangetype2} \times (\text{qh} - \text{enq});
\text{end;}
\text{end; changetype2 if loop}
\text{end; changetype2 for loop}
\text{elseif r2 > M}
\%
\text{The second change occurs after a forager using either policy will discover that the qo patch has changed.}
\%
\text{The contribution to dif in this case is negligible}
\text{dif} = \text{dif} + 0;
\%
\text{end; r2 if loop}
\text{end; r2 for loop}
\text{elseif qualitytype1 == 2 %q'\textgreater qh}
\text{probqualitytype1} = (1-\lambda)^{\text{qh}};
\%
\text{In this case the qo patch has changed to something worth being at, this is the case that motivates checking in on the qo quality patch.}
\%
\text{The expected new quality (enq) of the formerly qo quality}
enq = (qh + 1/lambda);

% We next need to consider when the second change occurs
for r2 = 1:M
  if r2 ≤ m
    probr2 = gamma^(r2-1)*(1-gamma);
  elseif r2 < M
    probr2 = gamma^m * delta^(r2-m-1) * (1-delta);
  elseif r2 == M
    % we let r2 == M represent all r2 values greater than
    % or equal to M
    probr2 = gamma^m * delta^(M-m-1);
  end;
  if r2 < m
    % In this case the contribution to dif is negligible
    dif = dif + 0;
  elseif r2 == m
    % We then need to consider what this second change is
    for changetype2 = 1:2
      if changetype2 == 1
        % The change is the session ending
        % probchangetype2 = Pd/(Pd + Pc/2 + Pd*(Pc/2));
        % In this case there is no contribution to dif
        dif = dif + 0;
      elseif changetype2 == 2
        % The change is the qh quality patch changing
        probchangetype2 = Pc/2*(1-Pd)... / (Pd + Pc/2 + Pd*(Pc/2));
        if rl1 < rl2
          dif = dif + probr1 * probchangetype1...
               * probqualitytype1 * probr2...
               * probchangetype2 * (enq - 1/lambda);
        elseif rl2 < rl1
          dif = dif + probr1 * probchangetype1...
               * probqualitytype1 * probr2...
               * probchangetype2 * (1/lambda - enq);
        end;
      end;
    end; % changetype2 if loop
end; %changetype2 for loop

elseif and(m < r2, r2 < M)
%We then need to consider the type of this second
%change.
for changetype2 = 1:3
if changetype2 == 1
%The change is the session ending
probchangetype2 = Pd/(Pd + Pc - Pd*Pc);

if rl1 < rl2
dif = dif + probr1 * probchangetype1...
   * probqualitytype1 * probr2...
   * probchangetype2 * (enq-qh)*(r2-m);
elseif rl2 < rl1
dif = dif + probr1 * probchangetype1...
   * probqualitytype1 * probr2...
   * probchangetype2 * (qh-enq)*(r2-m);
end;
elseif changetype2 == 2
%The change is the qh quality patch changing
probchangetype2 = (Pc/2)*(1-Pd)/(Pd + Pc - Pd*Pc);

if rl1 < rl2
dif = dif + probr1 * probchangetype1...
   * probqualitytype1 * probr2...
   * probchangetype2...
   * ((enq-qh)*(r2-m) + (enq - 1/lambda));
elseif rl2 < rl1
dif = dif + probr1 * probchangetype1...
   * probqualitytype1 * probr2...
   * probchangetype2...
   * ((qh-enq)*(r2-m) + (1/lambda - enq));
end;
elseif changetype2 == 3
%The change is the qo quality patch changing
probchangetype2 = (Pc/2)*(1-Pd)/(Pd + Pc - Pd*Pc);
if r11 < r12
    dif = dif + probr1 * probchangetype1...
    * probqualitytype1 * probr2...
    * probchangetype2...
    * ((enq-qh)*(r2-m) + (1/lambda - qh));
elseif r12 < r11
    dif = dif + probr1 * probchangetype1...
    * probqualitytype1 * probr2...
    * probchangetype2...
    * ((qh-enq)*(r2-m) + (qh - 1/lambda));
end;
end; %changetype2 if loop
end; %changetype2 for loop

elseif r2 >= M
    %The type of the change is not so important since both
    %foragers will be at the same patch M rounds after the
    %first change.

    if r11 < r12
        dif = dif + probr1 * probchangetype1...
        * probqualitytype1 * probr2...
        * probchangetype2 * (enq-qh)*(M-m);
    elseif r12 < r11
        dif = dif + probr1 * probchangetype1...
        * probqualitytype1 * probr2...
        * probchangetype2 * (qh-enq)*(M-m);
    end;
    end; %r2 if loop
end; %r2 for loop

end; %qualitytype1 if loop
end; %qualitytype1 for loop

end; %changetype1 if loop
end; %changetype1 for loop

end; %r1 for loop

end %end of function
Using the function *incredif* defined above this next piece of code computes an approximation of the optimal $t_\ast^{*}(q_h, q_o)$ for all of the patch quality pairs below a certain threshold.

%This script computes the optimal foraging strategy in the simplest %possible case. There are two patches, x and y, and a lone forager.

%Each round the foraging session ends with probability Pd
Pd = 0.02;

%Each round a patch changes with probability pc, with each patch being %equally likely to change, and its quality is redrawn from the quality %distribution.
Pc = 0.1;

%The probability of no change occurring is delta.
delta = (1-Pd)*(1-Pc);

%The probability of the foragers current patch not changing and the session %not ending is gamma
gamma = (1-Pd)*(1-Pc);

%If a patch changes quality its new quality is redrawn from a geometric %distribution with parameter lambda.
lambda = 0.1;

%There is no explicit cost to switching, so on each round a forager must %simply choose whether it would like to forage at the x patch or the y %patch.
n = 1/lambda * 6;

%The optimal strategy consists simply of how frequently the duck should %check in on the other patch to see if it has improved.

%For some values of qh and qo it may never be worthwhile to check in on the %qo quality patch, if $T_c(q_h, q_o) >$ cutoff, then we assume that it is never %worth checking in on the

prob_cutoff_wrong = 0.001;

cutoff = ceil(log(prob_cutoff_wrong)/log(gamma));
Tc = ones(n);

for qo = 1:n
    qo
    for qh = qo:n
        qh
        if qh == qo
            %if the patches have equal value switch back and forth between
            %them every round since you’ve got nothing to lose.
            Tc(qh,qo) = 1;
        else
            %Since Tc must be increasing in qh we start t at what it was
            %for the last value of qh.
            t = Tc(qh-1,qo);

            if t <= cutoff
                while incredif(t+1,qh,qo,Pd,Pc,lambda) > 0
                    t=t+1;
                    if t > cutoff
                        %In this case it looks like it might never be a
                        %good idea to check in on the low quality patch, in
                        %that case we stop and move on.
                        break
                    end;
                end;
            end;
            Tc(qh,qo) = t;
        end;
    end;
end;

B.2 Program for computing $d^*$ and $s^*$

First we have a computer program for computing $D_{ss}(q_h, q_o, d^*, s^*)$.

function diffs = stayswitchdifferences(x,qh,qo,lambda,Pc,Pd)
    d = x(1);
    s = x(2);
\textbf{APPENDIX B. COMPUTER PROGRAMS}

\texttt{Pcah = Pc/(2 - Pc * ( 1 - lambda * (1-lambda)^(qh-1) ));}

\texttt{Pcao = Pc/(2 - Pc * ( 1 - lambda * (1-lambda)^(qo-1) ));}

\texttt{hdiff = (1-Pcah) * ( (1-Pc) * (qh/d - qo/(1-d))...}
\texttt{+ Pc/2 * Dcn(d,s,qh,lambda,Pd)...}
\texttt{+ Pc/2 * -1 * Dcn(1-d,s,qo,lambda,Pd))...}
\texttt{+ Pcah * Dcp(d,s,qh,lambda,Pd);}

\texttt{odiff = (1-Pcao) * ( (1-Pc) * (qo/(1-d) - qh/d)...}
\texttt{+ Pc/2 * Dcn(1-d,s,qo,lambda,Pd)...}
\texttt{+ Pc/2 * -1 * Dcn(d,s,qh,lambda,Pd))...}
\texttt{+ Pcao * Dcp(d,s,qh,lambda,Pd);}

\texttt{diffs = [hdiff, odiff];}

Next we have a computer program for computing $D_{ho}(q_h, q_o, d^*, s^*)$

\texttt{function diff = Dho(d,s,qh,qo,lambda,Pc,Pd)
  diff = (1-Pc) * (qh/d - qo/(1-d))...}
\texttt{+ Pc/2 * (Dcn(d,s,qh,lambda,Pd))...}
\texttt{- Pc/2 * (Dcn(1-d,s,qo,lambda,Pd));
end}

Both of these programs make use of the functions $D_{cn}$ and $D_{cp}$, which are computed by the following programs.

\texttt{function diff = Dcn(d,s,q,lambda,Pd)
  \%This function computes the difference in expected lifetime payoff between
  \%a forager that happens to be a q quality patch, with d other foragers,
  \%when a formerly qo quality patch changes, and when there will be s
  \%foragers switching from the q patch to the changed patch on the next
  \%round.
  \%
  \%We need a parameter for how close we let certain values get to zero.
  epsilon = 0.00001;
  \%
  \%If there are as many switchers as foragers at one of the patch, we put a
  \%little buffer of epsilon foragers remaining at each patch to avoid zero
  \%divides.
  if d-s < epsilon

s = d-epsilon;
elseif 1-d-s < epsilon
    s = 1-d-epsilon;
end;

if s < epsilon
    s = epsilon;
end;

%If the qo quality patch changes to something with value less than or equal
to Qb then there will be too many switchers showing up there for
%equilibrium to be achieved.
if s == epsilon
    QB = 0;
else
    QB = floor((q * s) / (1-s));
end;

%If the qo quality patch changes to something of value greater than or
equal to QB then there will be too few switchers showing up there for
%equilibrium to be achieved

if s == epsilon
    QB = q+1;
elseif (1-d) + s < 1 - epsilon
    %A drastic improvement is only possible if something less than all of
    %the foragers are showing up at the changed patch
    QB = ceil((q * ((1-d) + s)) / ...
        (1 - ((1-d) + s) ));
else
    QB = -1;
    %This indicates that a drastic improvement is not possible since all of
    %the foragers are showing up at the changed patch
end;

%If the qo quality patch improves to the point where there are too few
%foragers showing up, it may or may not be the case that on the round after
%there will be too many foragers showing up to see how good this improved
%patch really is.
if QB == -1
    QBB = -1;
else
    QBB = ceil((q * (d-s))/ ... ( 1 - (d-s) ));
end;

%Then we compute the probability that qo' will fall into
%one of four categories qo'<=Qb, Qb<qo'<QB, QB <=qo'<QBB,
%and QBB <= qo' while still QB <= qo'.

%The probability that qo' <= Qb is denoted
pQb = 1-(1-lambda)^(Qb);

%The probability that Qb < qo' < QB is denoted
%pEQ = (1-lambda)^(Qb) - (1-lambda)^(QB-1);

%The probability that QB <= qo' is denoted
if QB == -1
    pQB = 0;
else
    pQB = (1-lambda)^(QB-1);
end;

%The probability that QB <= qo' < QBB is
if QB < QBB
    pQBB = (1-lambda)^(QB-1) - (1-lambda)^(QBB-1);
else
    pQBB = 0;
end;

%Then we compute the expected value of qo' given that it
%falls into one of three categories, qo'<= Qb, QB<=qo', and
%QB <= qo' < QBB. Denoted littlemean, bigmean, and middle
%mean respectively
if Qb > 0
    index = 1:Qb;
    littlemean = (index * ((1-lambda).^(index-1))') / ... sum((1-lambda).^(index-1));
else
    littlemean = 0;
end;

if QB == -1
    bigmean = 0;
else
    bigmean = QB + 1/lambda - 1;
end;

if QBB <= QB
    middlemean = 0;
else
    index = QB:(QBB-1);
    middlemean = (index * ((1-lambda).^(index-1))') / ... 
                sum((1-lambda).^(index-1));
end;

diff = ( q/d - 1/lambda/(1-d) ) ... 
      + (1-Pd) * pQb * (s/d) * (littlemean/s - q/(1-s)) ... 
      + (1-Pd) * pQB * (1 - s/d) * (q/(d-s) - bigmean/(d + s)) ... 
      + (1-Pd)^2 * pQBB * (1-s/d) * (middlemean/(d-s) - q/(1-d+s));
end

function diff = Dcp(d,s,q,lambda,Pd)
    epsilon = 0.00001;

    if d-s < epsilon
        s = d-epsilon;
    elseif 1-d-s < epsilon
        s = 1-d-epsilon;
    end;

    if s < epsilon
s = epsilon;
end;

%If the qo quality patch changes to something with value less than or equal 
%to Qb then there will be too many switchers showing up there for 
%equilibrium to be achieved.
if s == epsilon
    Qb = 0;
else
    Qb = floor((q * s) / (1-s));
end;

%If the qo quality patch changes to something of value greater than or 
%equal to QB then there will be too few switchers showing up there for 
%equilibrium to be achieved.
if s == epsilon
    QB = q+1;
elseif (1-d) + s < 1 - epsilon
    %A drastic improvement is only possible if something less than all of 
    %the foragers are showing up at the changed patch
    QB = ceil((q * ((1-d) + s)) / ... 
        (1 - ((1-d) + s)));
else
    QB = -1;
    %This indicates that a drastic improvement is not possible since all of 
    %the foragers are showing up at the changed patch
end;

%If the qo quality patch improves to the point where there are too few 
%foragers showing up, it may or may not be the case that on the round after 
%there will be too many foragers showing up to see how good this improved 
%patch really is.
if QB == -1
    QBB = -1;
else
    QBB = ceil((q * (d-s)) / ... 
        (1 - (d-s)));
end;

%Then we compute the probability that qo’ will fall into 
%one of four categories qo’<=Qb, Qb<qo’<QB, QB <=qo’<QBB,
%and QBB <= qo’ while still QB <= qo’.

%The probability that qo’ <= Qb is denoted
pQb = 1-(1-lambda)^(Qb);

%The probability that Qb < qo’ < QB is denoted
%pEQ = (1-lambda)^(Qb) - (1-lambda)^(QB-1);

%The probability that QB <= qo’ is denoted
if QB == -1
   pQB = 0;
else
   pQB = (1-lambda)^(QB-1);
end;

%The probability that QB <= qo’ < QBB is
if QB < QBB
   pQBB = (1-lambda)^(QB-1) - (1-lambda)^(QBB-1);
else
   pQBB = 0;
end;

%Then we compute the expected value of qo’ given that it
%falls into one of three categories, qo’<= Qb, QB<=qo’, and
%QB <= qo’ < QBB. Denoted littlemean, bigmean, and middle
%mean respectively
if Qb > 0
   index = 1:Qb;
   littlemean = (index * ((1-lambda).^(index-1))') / ... 
   sum((1-lambda).^(index-1));
else
   littlemean = 0;
end;

if QB == -1
   bigmean = 0;
else
   bigmean = QB + 1/lambda - 1;
end;
if QBB <= QB
    middlemean = 0;
else
    index = QB:(QBB-1);
    middlemean = (index * ((1-lambda).^(index-1)))' / ... 
                 sum((1-lambda).^(index-1));
end;

diff = pQb * (q/(1-s) - littlemean/s) ... 
      + pQB * (q/(d-s) - bigmean/(d + s)) ... 
      + (1-Pd) * pQBB * (middlemean/(d-s) - q/(1-d+s));
end

Using these functions the following program computes when possible the \( d, s \) pair such that \( D_{ss}(q_h, q_o, d, s) = D_{ss}(q_o, q_h, 1 - d, s) = 0 \) and if this is not possible it finds the \( (d, s) \) pair such that \( |D_{ss}(q_h, q_o, d, s)| + |D_{ss}(q_o, q_h, 1 - d, s)| \) is minimized while \( D_{ho}(q_h, q_o, d, s) = 0 \).

function [EQD, EQS, QC] = freeinfoeq(deathprob,changeprob,qualitydist)
%This function computes the ES foraging strategy (or the closest thing to %an ES strategy we can find) in the simplest possible case. There are two %patches, x and y, and an infinite population of foragers.

%Each round the foraging session ends with probability \( pd \)
Pd = deathprob; %usually 0.02

%Each round there is a chance that one of the patches will change quality %with probability \( pc \).
Pc = changeprob; %usually 0.1

%If there is a patch quality change then only one of the patches changes %quality, and each patch is equally likely to change quality regardless of %its current quality and the proportion of the population foraging there. %If a patch changes quality its new quality is redrawn from a geometric %distribution with parameter \( \lambda \) and mean \( 1/\lambda \).
lambda = qualitydist; %usually 0.1

%There is no explicit cost to switching, so on each round a forager must %simply choose whether it would like to forage at the x patch or the y
%patch.
n = 1/\lambda * 4;

%If differences between two things are less than epsilon we count them as
%being the same thing
epsilon = 0.001;

%The ES strategy can be described by the equilibrium proportion of foragers
%per patch for each possible pair of patch qualities. And the proportion of
%the population that will switch back and forth between these
%patches.

%We store the equilibrium distribution of foragers in a matrix EQD such
%that EQD(qx,qy) gives the equilibrium proportion of foragers at the x
%patch.

EQD = zeros(n);
maxEQD = EQD;
minEQD = EQD;
guessEQD = EQD;
%A good initial guess for EQD is the value such that the expected quality
%per forager on the next round is equal at both patches
for qh = 1:n
  for qo = 1:n
    guessEQD(qh,qo) = (qh*(1-Pc/2) + Pc/2/\lambda) / ...
      ( Pc/\lambda + (qh+qo) * (1-Pc/2));
  end;
end;
%We store the equilibrium switching regime of these foragers in a matrix
%EQS such that EQS(qx,qy) gives the proportion of the total population that
%will switch from the qx quality patch to the qy quality patch and vice
%versa at equilibrium.

EQS = zeros(n);
%A good initial guess of EQS is ...
for qh = 1:n
  for qo = 1:n
    EQS(qh,qo) = (min(guessEQD(qh,qo),guessEQD(qo,qh)))/2;
  end;
end;
% We keep track of whether a pair of values is such that a maximal switching 
% policy is the evolutionarily singular point \( QC(q_h, q_o) = 2 \), \( QC \) for quality 
% class, or if some intermediate switching policy is the evolutionarily 
% singular point.

\[ QC = \text{zeros}(n); \]

% In population using a policy of the standard form, with \( EQD \) as their 
% target population distribution and \( EQS \) as their switching regime, when a 
% change occurs at one of the patches one of four things can happen. 
% The change can be a drastic worsening, a manageable change, a drastic 
% improvement, and a drastic improvement with a twist.

% We let \( Q_b(q_h, q_o) \) be the value for which if a \( q_o \) quality patch changes 
% to something worse than or equal to \( Q_b(q_h, q_o) \), given that the other patch 
% has quality \( q_h \) then equilibrium cannot be immediately achieved.

% We let \( QB(q_h, q_o) \) be the value for which if a \( q_o \) quality patch changes 
% to something better than or equal to \( QB(q_h, q_o) \), given that the other patch 
% has quality \( q_h \) then equilibrium cannot be immediately achieved.

% We let \( QBB(q_h, q_o) \) be the value for which if a \( q_o \) quality patch changes 
% to something better than \( QB(q_h, q_o) \), but not better than \( QBB(q_h, q_o) \), then 
% equilibrium will not be achieved immediately or the round after that.

for \( q_h = 1:n \)
    for \( q_o = q_h:n \)
        if \( q_h == q_o \)
            \( EQD(q_h, q_o) = 0.5; \)
            \( EQS(q_h, q_o) = 0.5; \)
        else
            intguess = \([\text{guessEQD}(q_h, q_o)/2; \text{EQS}(q_h, q_o)/2];\)
            \( f = @(x)\text{stayswitchdifferences}(x, q_h, q_o, \lambda, P_c, P_d); \)
            yone = fsolve\( f, \text{intguess}; \)
            if and(and(and(yone(1) > yone(2), yone(1) > 0), yone(2) > 0), yone(1) < 1)
In this case it appears that a valid solution has been found and so this (d,s) combo is a candidate ES policy point.

First we see if we can make that solution just a little more valid.
\[
\text{checkone} = \text{sum}(\text{abs}((\text{stayswitchdifferences}(\ldots
\text{yone},\text{qh},\text{qo},\lambda,P_c,P_d))))
\]
if checkone > 0.00001
\[
\text{yone} = \text{fsolve}(f,\text{yone})
\]
\[
\text{checktwo} = \text{sum}(\text{abs}((\text{stayswitchdifferences}(\ldots
\text{yone},\text{qh},\text{qo},\lambda,P_c,P_d))))
\]
end;

QC(qh,qo) = 1;
EQD(qh,qo) = yone(1);
EQD(qh,qo) = 1-yone(1);
EQS(qh,qo) = yone(2);
EQS(qh,qo) = yone(2);
if Dho(yone(1),yone(2),1,2,\lambda,P_c,P_d) > 0.000001
  'just curious'
end;

else
In the case where one cannot find a (d,s) pair that makes switching versus staying neutral we then try to find the combination of s and d that minimizes the incentive to switch. Interestingly this is always the same value as if s is set to be min(d,1-d) and the value of d is found such that the value of being at each patch given that the neither patch has changed on the current round is equal.

f = @(d)maxeqdiff(d,qh,qo,\lambda,P_c,P_d);
maxeqd = fsolve(f,guessEQD(qh,qo));

maxdiffs = stayswitchdifferences(...
  [maxeqd,maxeqd],qh,qo,\lambda,P_c,P_d);

f = @(x)ssdiffstotal(x,qh,qo,\lambda,P_c,P_d);
ytwo = fmincon(...
f, intguess, [-1, 1], 0, [], [], [0.01, 0.01], [0.5, 0.5]);

mindiffs = stayswitchdifferences(ytwo, qh, qo, lambda, Pc, Pd);

% Also just to be safe we use this as a guess for the % solving approach
f = @(x) stayswitchdifferences(x, qh, qo, lambda, Pc, Pd);
ythree = fsolve(f, ytwo);

if and(and(and(ythree(1) > ythree(2), ythree(1) > 0), ... ythree(2) > 0), ythree(1) < 1)

    QC(qh, qo) = 1;
    EQD(qh, qo) = ythree(1);
    EQD(qo, qh) = 1 - ythree(1);
    EQS(qh, qo) = ythree(2);
    EQS(qo, qh) = ythree(2);

else
    if abs(maxeqd - ytwo(1)) < epsilon
        EQD(qh, qo) = maxeqd;
        EQD(qo, qh) = 1 - maxeqd;
        EQS(qh, qo) = maxeqd;
        EQS(qo, qh) = maxeqd;
    else
        if sum(abs(maxdiffs)) <= sum(abs(mindiffs))
            EQD(qh, qo) = maxeqd;
            EQD(qo, qh) = 1 - maxeqd;
            EQS(qh, qo) = maxeqd;
            EQS(qo, qh) = maxeqd;
        else
            QC(qh, qo) = 2;
            EQD(qh, qo) = ytwo(1);
            EQD(qo, qh) = 1 - ytwo(1);
            EQS(qh, qo) = ytwo(2);
            EQS(qo, qh) = ytwo(2);
        end;
    end;
end;
eon;
```matlab
function totaldiff = ssdiffstotal(x,qh,qo,lambda,Pc,Pd)
    d = x(1);
    s = x(2);

    Pcah = Pc/(2 - Pc * (1 - lambda * (1-lambda)^((qh-1))));
    Pcao = Pc/(2 - Pc * (1 - lambda * (1-lambda)^((qo-1))));

    hdiff = (1-Pcah) * ((1-Pc) * (qh/d - qo/(1-d)) + Pc/2 * Dcn(d,s,qh,lambda,Pd) + Pc/2 * -1 * Dcn(1-d,s,qo,lambda,Pd)) + Pcah * Dcp(d,s,qh,lambda,Pd);

    odiff = (1-Pcao) * ((1-Pc) * (qo/(1-d) - qh/d) + Pc/2 * Dcn(1-d,s,qo,lambda,Pd) + Pc/2 * -1 * Dcn(d,s,qh,lambda,Pd)) + Pcao * Dcp(d,s,qh,lambda,Pd);

    totaldiff = abs(hdiff) + abs(odiff);
end

function diffs = stayswitchdifferences(x,qh,qo,lambda,Pc,Pd)
    d = x(1);
    s = x(2);

    Pcah = Pc/(2 - Pc * (1 - lambda * (1-lambda)^((qh-1))));
    Pcao = Pc/(2 - Pc * (1 - lambda * (1-lambda)^((qo-1))));

    hdiff = (1-Pcah) * ((1-Pc) * (qh/d - qo/(1-d)) + Pc/2 * Dcn(d,s,qh,lambda,Pd) + Pc/2 * -1 * Dcn(1-d,s,qo,lambda,Pd)) + Pcah * Dcp(d,s,qh,lambda,Pd);
```

odiff = (1-Pcao) * ( (1-Pc) * (qo/(1-d) - qh/d)...
+ Pc/2 * Dcn(1-d,s,qo,lambda,Pd)...
+ Pc/2 * -1 * Dcn(d,s,qh,lambda,Pd))...
+ Pcao * Dcp(d,s,qh,lambda,Pd);

diffs = [hdiff, odiff];
end

B.3 Program for computing optimal policy in problem 4.1

The following program uses expressions 84, 85 and 87, to compute the whether it is better to observe, exploit or switch on the first round for a given initial patch quality $q_x$, and given problem parameters.

function [payoffexploit payoffobserve payoffswitch] =...
duckproblemzero(initialvalue, payoffdistribution, timealive, ...
costobserve, costswitch)

l = timealive;

%The payoff distribution is geometric with parameter p and mean 1/p
p = payoffdistribution;

%There are two patches and this is the payoff at the patch you are at. You %don’t know the value at the other patch.
qx = initialvalue;

%When you do these other actions you receive a fraction of what you would %have recieved if you had been foraging.
co = costobserve;
cs = costswitch;

payoffexploit = qx*l;
payoffobserve = co*qx ...
+ (1-p)^(floor(qx*(l-1)/(l-2+cs)))*...
(1/p+floor(qx*(l-1)/(l-2+cs)))*(1-2+cs)...
+ (1-(1-p)^(floor(qx*(l-1)/(l-2+cs))))*qx*(l-1);
payoffswitch = cs*1/p...
Using the above program we can then find the critical values of $q_x$ for which it the optimal actions transitions from switching, to observing, to exploiting.

% This is a script that computes the threshold values for when a single duck % should exploit, observe, or switch, given the current value of their patch % and the payoff distribution parameter lambda, the duration of the foraging % session T and the costs of switching and observing.

lambda = 0.01; % parameter of the geometric payoff distribution

T = 100; % duration of the foraging session

co = 0.9; % observation cost factor

cs = 0.5; % switching cost factor

test = 1;
k = 0;

swtobsthreshold = -1;
obsexpthreshold = -1;

while test > 0
    k = k + 1; % the quality of the starting patch
    [payoffexploit payoffobserve payoffswitch] = ...
        duckproblemzero(k, lambda, T, co, cs);

    if swtobsthreshold == -1
        if payoffobserve > payoffswitch
            swtobsthreshold = k;
        end;
    elseif obsexpthreshold == -1
        if payoffexploit > payoffobserve
            obsexpthreshold = k;
        end;
    else

function [probS probO] = oneshotes(qx)

% This function computes the ES strategy for a population of foragers who all
% initially start at the x patch with value qx, and who do not know the
% value of the y patch qy, but they know it is from a geometric
% distribution with parameter
lambda = 0.1;
% They know that the foraging session lasts for T rounds
T = 50;
% The cost factor of observing is
co = 0.9;
% The cost factor of switching is
cs = 0.5;

% We need to set the probability threshold at which we truncate the
% payoff distribution. Cases that have this probability or less of happening
% are not considered.
cutoff = 0.00000001;

% beta is a parameter of the replicator dynamics, it determines the rate of

B.4 Implementation of Replicator Dynamics and Dynamic Programming

The following computer program uses dynamic programing to compute the relative
values of each action in each belief state given that the population is using a particular policy, and then uses a simple replicator dynamic to increase the probability with which actions with higher values are chosen and decrease the probability with which actions with lower values are chosen in the residents populations policy. The value of each action in each belief state is then recomputed using this new policy, and again the policy is updated using a simple replicator dynamic. This process is repeated until eventually a policy is found such that the value of each action that is engaged in with some positive probability in each belief state is equal.
%change of the probabilities of choosing each action, on each iteration.
\[ \beta = \frac{1}{16}; \]

%We need to set a threshold for how little the strategy changes in order to
%count as stabilized.
\[ \text{strategyepsilon} = 0.00005; \]

%We need to initialize the population strategy. We know that observing and
%uniformed switching only happens on the odd numbered rounds. \( \text{probS}(t) \) and
%\( \text{probO}(t) \) are the probabilities with which a forager at the x patch
%switches and observes on a round when the x patch is not at equilibrium.
\[
\text{probS} = \text{zeros}(1,T);
\text{for } t = 1:2:T
\quad \text{probS}(t) = \frac{1}{3};
\text{end;}
\text{probO} = \text{probS};
\]

%If you have a better initial guess load it here
\load{savetime}
\text{newprobS} = \text{probS};
\text{newprobO} = \text{probO};
\text{probE} = 1 - \text{probO} - \text{probS};
\text{newprobE} = \text{probE};

\text{test} = 1;
\text{while } \text{test} > \text{strategyepsilon}
\quad \text{count} = \text{count} + 1;
\quad \%Vx andVy are arrays containing the value to go of each state, so
\quad \%Vx(equilibriumtype, round) gives the value to go for being at the x
\quad \%patch, on a given round, and the equilibrium type. Round goes from 1
\quad \%to T. Equilibrium type 1 is when too many foragers switch from the x
\quad \%patch to the y patch and either some of them switch back, or foragers
\quad \%already at the y patch preemptively switch back to the x patch.
\quad \%Equilibrium type 2 is when just the right amount of foragers switch so
\quad \%that there is no incentive for further switching. Equilibrium type 3
\quad \%is when too few foragers switch to the patch and so the equilibrium is
\quad \%brought about by observers at the x patch switching to the y patch.
\quad \%Equilibrium type 4 is when the population distribution is no yet at
\quad \%equilibrium. The cases are organized in terms of which occurs as qy
%increases.
Vx = zeros(4,T);
Vy = zeros(4,T);

%Vactions keeps track of the expected value resulting from a given
%action on a given round. Since decisions are only made on odd rounds
%Vactions is only recorded on odd rounds.
Vactions = zeros(3,T);

%Caseprobs gives the probability of being in each type of equilibrium
%contingent on not being in equilibrium on the even round prior. Note
%that only the odd entries of case probs have non-zero entries.
caseprobs = zeros(4,T);

%EQY gives the expected quality at the y patch for a forager at the x
%patch given that the population is not at equilibrium on the even
%round prior and given the type of equilibrium that occurs in the
%subsequent rounds.
EQY = zeros(4,T);

%DY and DX are vectors denoting the proportion of the population at the
%x and y patches respectively if equilibrium has not been achieved.
DX = zeros(1,T);
DY = DX;
for t = 1:T
    if t == 1
        DY(t) = probS(1);
        DX(t) = 1-DY(t);
    elseif t == 2
        DY(t) = probS(1) + probO(1);
        DX(t) = 1 - DY(t);
    else
        DY(t) = DY(t-1) + DX(t-1)*probS(t) + DX(t-2)*probO(t-1);
        DX(t) = 1 - DY(t);
    end;
end;

offset = 1 - (1-cs)./(T-(0:T-1));
for t = 1:T
    if t == 1
baseline = 0;

crit1 = floor(qx * DY(1)/DX(1) * offset(2));
crit2 = floor(qx * DY(1)/DX(1) / offset(2));
crit3 = floor(qx * DY(2)/DX(2) / offset(2));
elseif mod(t,2) == 1
  baseline = qx * DY(t-1)/DX(t-1) / offset(t-1);
crit1 = floor(qx * DY(t)/DX(t) * offset(t))-floor(baseline);
crit2 = floor(qx * DY(t)/DX(t) / offset(t+1))-floor(baseline);
crit3 = floor(qx * DY(t+1)/DX(t+1) / ... 
  offset(t+1))-floor(baseline);
elseif mod(t,2) == 0
  \%don’t do a thing sweet heart
end;

caseprobs(1,t) = geocdf(crit1-1,lambda);
caseprobs(2,t) = geocdf(crit2-1,lambda) - geocdf(crit1-1,lambda);
caseprobs(3,t) = geocdf(crit3-1,lambda) - geocdf(crit2-1,lambda);
caseprobs(4,t) = 1 - geocdf(floor(crit3)-1,lambda);

if caseprobs(1,t) > 0
  EQY(1,t) = geopdf(0:crit1-1,lambda) * ...
    (1:crit1)'/caseprobs(1,t) + floor(baseline);
else
  EQY(1,t) = 0;
end;
if caseprobs(2,t) > 0
  EQY(2,t) = geopdf(crit1:crit2-1,lambda) * (crit1+1:crit2)'...
    /caseprobs(2,t) + floor(baseline);
else
  EQY(2,t) = 0;
end;
if caseprobs(3,t) > 0
  EQY(3,t) = geopdf(crit2:crit3-1,lambda) * (crit2+1:crit3)'...
    /caseprobs(3,t) + floor(baseline);
else
  EQY(3,t) = 0;
end;

EQY(4,t) = floor(crit3) + floor(baseline) + 1/lambda;

% Now that we have all these values we can compute the value of the % equilibrium states
if t == 1
  % Value of being at each patch when a type 1 equilibrium occurs % on round 2.
  Vx(1,t+1) = (EQY(1,t)/offset(t+1) + qx)*(T-t-1);
  Vy(1,t+1) = (EQY(1,t) + qx*offset(t+1))*(T-t-1);

  % This is for debugging purposes
  if abs(Vx(1,t+1)*(1+cs/(T-t-1))-Vy(1,t+1)*(1+1/(T-t-1)))... > 10^-9
    'oh no'
  end;

% Value of being at each patch when a type 2 equilibrium occurs % on round 1.
Vx(2,t) = qx/DX(t)*(T-t);
Vy(2,t) = EQY(2,t)/DY(t)*(T-t);

% Value of being at each patch when a type 3 equilibrium occurs % on round 2.
Vx(3,t+1) = (qx + EQY(3,t)*offset(t+1))*(T-t-1);
Vy(3,t+1) = (qx/offset(t+1) + EQY(3,t))*(T-t-1);

% More debugging
if abs(Vx(3,t+1)*(1+1/(T-t-1))-Vy(3,t+1)*(1+cs/(T-t-1)))... > 10^-9
  'oh no'
end;
elseif t == T-1
  if mod(t,2) == 1
    % value of being at each patch when a type 1 equilibrium % occurs on the round before the last round.
\[ Vx(1,t) = \frac{EQY(1,t)}{offset(t)} + qx \times (T-t); \]
\[ Vy(1,t) = (EQY(1,t) + qx \times offset(t)) \times (T-t); \]

% value of being at each patch when a type 2 equilibrium
% occurs on the round before the last round.
\[ Vx(2,t) = qx / DX(t) \times (T-t); \]
\[ Vy(2,t) = EQY(2,t) / DY(t) \times (T-t); \]

% value of being at each patch when a type 3 equilibrium
% occurs on the last round. Even though this should be zero
% for computational ease, we set it to the per round payoff.
\[ Vx(3,t+1) = 0; \]
\[ Vy(3,t+1) = 0; \]
end;
elseif t == T
% everything is already zero as it should be
elseif mod(t,2) == 1
% value of being at each patch when a type 1 equilibrium occurs
% on round \( t \).
\[ Vx(1,t) = \frac{EQY(1,t)}{offset(t)} + qx \times (T-t); \]
\[ Vy(1,t) = (EQY(1,t) + qx \times offset(t)) \times (T-t); \]

if abs(Vx(1,t)*(1+cs/(T-t))-Vy(1,t)*(1+1/(T-t))) > 10^{-9}
    'oh no'
end;

% value of being at each patch when a type 2 equilibrium occurs
% on round \( t \).
\[ Vx(2,t) = qx / DX(t) \times (T-t); \]
\[ Vy(2,t) = EQY(2,t) / DY(t) \times (T-t); \]

% value of being at each patch when a type 3 equilibrium occurs
% on round \( t+1 \).
\[ Vx(3,t+1) = (qx + EQY(3,t) \times offset(t+1)) \times (T-t-1); \]
\[ Vy(3,t+1) = (qx / offset(t+1) + EQY(3,t)) \times (T-t-1); \]

if abs(Vx(3,t+1)*(1+1/(T-t-1))-Vy(3,t+1)*(1+cs/(T-t-1))) > 10^{-9}
    'oh no'
end;
end;

Ttest = prod(caseprobs(4,1:2:t));
if and(Ttest < cutoff, mod(t,2) == 0)
    % We can compute the decision rule up to this point
    releventT = t;
    break
end;
end;

% Now that we have the value at the terminal states we are able to crawl
% back from the terminal states and determine the value of each
% preceding state. We compute the value of a given state assuming that
% the given population strategy PS and PO is being employed.

for i = 0:releventT-1
    % We are computing the value for the non-terminal states of time t,
    % which depend only on the value of subsequent states. Note that we
    % are skipping over the value of states with even round number since
    % the behaviour of a forager is in a sense determined by what
    % they do on the odd numbered rounds.
    t = releventT-i;

    % Compute the probability with which each action is chosen
    if t == 1
        % If you're at the x patch and you exploit you get this on
        % average
        Vexp = ...
        caseprobs(1,t)* ( Vx(1,t+1)*(1+1/(T-t-1)) + qx/DX(t) ) ... 
        + caseprobs(2,t)* ( Vx(2,t)*(1+1/(T-t)) ) ...
        + caseprobs(3,t)* ( Vx(3,t+1)*(1+1/(T-t-1)) + qx/DX(t) ) ...
        + caseprobs(4,t)* ( Vx(4,t+1) + qx/DX(t) + qx/DX(t+1) );
    % If you're at the x patch and you observe you get this on
    % average
        Vobs = ...
        caseprobs(1,t)* ( Vx(1,t+1)*(1+1/(T-t-1)) + co*qx/DX(t) ) ...
        + caseprobs(2,t)* ( Vx(2,t)*(1+co/(T-t)) ) ...
        + caseprobs(3,t)* ( Vx(3,t+1)*(1+1/(T-t-1)) + co*qx/DX(t) ) ...
    end;
end;
APPENDIX B. COMPUTER PROGRAMS

+ caseprobs(4,t)*( Vy(4,t+1) + co*qx/DX(t) + ... 
  cs*EQY(4,t)/DY(t+1));
%If you’re at the x patch and you switch you get this on 
%average 
Vswt = ...
caseprobs(1,t)* ( Vx(1,t+1)*(1+cs/(T-t-1)) + ... 
  cs*EQY(1,t)/DY(t) ) ... 
+ caseprobs(2,t)*( Vy(2,t)*(1+cs/(T-t))) ... 
+ caseprobs(3,t)* ( Vy(3,t+1)*(1+1/(T-t-1)) + ... 
  cs*EQY(3,t)/DY(t)) ... 
+ caseprobs(4,t)* ( Vy(4,t+1) + cs*EQY(4,t)/DY(t) + ... 
  EQY(4,t)/DY(t+1));
%Given the strategy used, the value of being at the x patch on 
%average is.
Vavg = Vexp * probE(t) + Vobs * probO(t) + Vswt * probS(t); 
Vactions(:,t) = [Vexp;Vobs;Vswt];
%The population strategy is updated according to a replicator 
%dynamic so that actions with a higher payoff are used with 
%higher probability in the subsequent iteration.
newprobO(t) = probO(t)*(1 - beta*(Vavg-Vobs)/Vavg); 
newprobS(t) = probS(t)*(1 - beta*(Vavg-Vswt)/Vavg); 
newprobE(t) = 1 - newprobO(t) - newprobS(t);

elseif t == T-1 
%if this is the second to last round
if mod(t,2) == 1
  Vexp = ...
  caseprobs(1,t)* ( Vx(1,t)*(1+1/(T-t)) ) ... 
  + caseprobs(2,t)* ( Vx(2,t)*(1+1/(T-t)) ) ... 
  + caseprobs(3,t)* ( qx/DX(t) + (qx + EQY(3,t)*offset(t+1)) ) ... 
  + caseprobs(4,t)* ( qx/DX(t) + qx/DX(t+1) );

  Vobs = ...
  caseprobs(1,t)* ( Vx(1,t)*(1+co/(T-t)) ) ... 
  + caseprobs(2,t)* ( Vx(2,t)*(1+co/(T-t)) ) ... 
  + caseprobs(3,t)* ( co*qx/DX(t) + ... 
  cs*(qx/offset(t+1) + EQY(3,t)) ) ... 
  + caseprobs(4,t)* ( co*qx/DX(t) + cs*EQY(4,t)/DY(t+1) );

APPENDIX B. COMPUTER PROGRAMS

\[ \text{Vswt} = \text{caseprobs}(1,t) \times (\text{Vy}(1,t) \times (1 + cs/(T-t))) \ldots \\
+ \text{caseprobs}(2,t) \times (\text{Vy}(2,t) \times (1 + cs/(T-t))) \ldots \\
+ \text{caseprobs}(3,t) \times (\text{cs*EQY}(3,t)/\text{DY}(t) + \ldots \\
\text{qy/offset}(t+1) + \text{EQY}(3,t)) \ldots \\
+ \text{caseprobs}(4,t) \times (\text{cs*EQY}(4,t)/\text{DY}(t) + \text{EQY}(4,t)/\text{DY}(t+1)); \]

\[ \text{Vavg} = \text{Vexp} \times \text{probE}(t) + \text{Vobs} \times \text{probO}(t) + \text{Vswt} \times \text{probS}(t); \]
\[ \text{Vactions}(:,t) = [\text{Vexp};\text{Vobs};\text{Vswt}]; \]
\[ \text{Vx}(4,t-1) = \text{Vavg}; \]
\%Which is the value to go of being at the x patch when 
\%equilibrium has not been achieved by round t-1.

\%We have the value of each action, exploit, observe, switch. 
\%When the ES is employed all of these will have the same value, 
\%for now we shift the probabilities of using each action so 
\%that actions with high payoff are employed more frequently than 
\%actions with low payoff. We do this according to a simple 
\%replicator dynamic. The specifics of the dynamic are not 
\%important, we are just looking for the fixed point.

\[ \text{newprobO}(t) = \text{probO}(t) \times (1 - \beta \times (\text{Vavg} - \text{Vobs})/\text{Vavg}); \]
\[ \text{newprobS}(t) = \text{probS}(t) \times (1 - \beta \times (\text{Vavg} - \text{Vswt})/\text{Vavg}); \]
\[ \text{newprobE}(t) = 1 - \text{newprobO}(t) - \text{newprobS}(t); \]

\%This is the average value to go from being at the y patch not 
\%at equilibrium on round t-1.
\[ \text{Vy}(4,t-1) = \ldots \\
\text{caseprobs}(1,t) \times (\text{Vx}(1,t) \times (1 + cs/(T-t))) \ldots \\
+ \text{caseprobs}(2,t) \times (\text{Vx}(2,t) \times (1 + 1/(T-t))) \ldots \\
+ \text{caseprobs}(3,t) \times (\text{EQY}(3,t)/\text{DY}(t) \ldots \\
\text{qy/offset}(t+1) + \text{EQY}(3,t)) \ldots \\
+ \text{caseprobs}(4,t) \times (\text{EQY}(4,t)/\text{DY}(t) + \text{EQY}(4,t)/\text{DY}(t+1)); \]
end;

elseif t == T
\%if this is the last round
if mod(t,2) == 1;
\text{Vexp} = \text{qx}/\text{DX}(t);
Vobs = co*qx/DX(t);

Vswt = caseprobs(1,t)*( cs*EQY(1,T)/DY(t) )...  
    + caseprobs(2,t)* ( cs*EQY(1,T)/DY(t) )...   
    + caseprobs(3,t)* ( cs*EQY(3,t)/DY(t) ) ...   
    + caseprobs(4,t)* ( cs*EQY(4,t)/DY(t) );

Vavg = Vexp * probE(t) + Vobs * probO(t) + Vswt * probS(t);
Vactions(:,t) = [Vexp;Vobs;Vswt];
Vx(4,t-1) = Vavg;
   \%Which is the value to go of being at the x patch when
   \%equilibrium has not been achieved by round t-1.

\%We have the value of each action, exploit, observe, switch.
\%When the ES is employed all of these will have the same value,
\%for now we shift the probabilities of using each action so
\%that actions with high payoff are employed more freqently than
\%actions with low payoff. We do this according to a simple
\%replicator dynamic. The specifics of the dynamic are not
\%important, we are just looking for the fixed point.

newprobO(t) = probO(t)*(1 - beta*(Vavg-Vobs)/Vavg);
newprobS(t) = probS(t)*(1 - beta*(Vavg-Vswt)/Vavg);
newprobE(t) = 1 - newprobO(t) - newprobS(t);

\%This is the average value to go from being at the y patch not
\%at equilibrium on round t-1.
Vy(4,t-1) = ...  
    caseprobs(1,t)* ( EQY(1,T)/DY(t) )...  
    + caseprobs(2,t)* ( EQY(1,T)/DY(t) )...  
    + caseprobs(3,t)* ( EQY(3,t)/DY(t) ) ...  
    + caseprobs(4,t)* ( EQY(4,t)/DY(t) );

end;

elseif mod(t,2) == 1;
   \%recall the cases are:
   \%Case 1 too many foragers switch, equilibrium type 1 achieved
   \%Case 2 the right amount of foragers switch, equilibrium type 2
   \%achieved
Case 3 too few foragers switch, but there are enough observers to make equilibrium happen on the next round, equilibrium type 3 achieved.

Case 4 too few foragers switch and too few observers, equilibrium is not achieved.

If you’re at the x patch and you exploit you get this on average

\[ \text{Vexp} = \ldots \]

\[ \text{caseprobs(1,t)}*( \text{Vx}(1,t)*(1+1/(T-t)) ) \ldots + \text{caseprobs(2,t)}*( \text{Vx}(2,t)*(1+1/(T-t)) ) \ldots + \text{caseprobs(3,t)}*( \text{Vx}(3,t+1)*(1+1/(T-t-1)) + \text{qx}/\text{DX}(t) ) \ldots + \text{caseprobs(4,t)}*( \text{Vx}(4,t+1) + \text{qx}/\text{DX}(t) + \text{qx}/\text{DX}(t+1) ); \]

If you’re at the x patch and you observe you get this on average

\[ \text{Vobs} = \ldots \]

\[ \text{caseprobs(1,t)}*( \text{Vx}(1,t)*(1+\text{co}/(T-t)) ) \ldots + \text{caseprobs(2,t)}*( \text{Vx}(2,t)*(1+\text{co}/(T-t)) ) \ldots + \text{caseprobs(3,t)}*( \text{Vx}(3,t+1)*(1+\text{cs}/(T-t-1)) + \text{co*qx}/\text{DX}(t) ) \ldots + \text{caseprobs(4,t)}*( \text{Vy}(4,t+1) + \ldots \text{co*qx}/\text{DX}(t) + \text{cs*EQY(4,t)}/\text{DY}(t+1) ); \]

If you’re at the x patch and you switch you get this on average

\[ \text{Vswt} = \text{caseprobs(1,t)}*( \text{Vy}(1,t)*(1+\text{cs}/(T-t)) ) \ldots + \text{caseprobs(2,t)}*( \text{Vy}(2,t)*(1+\text{cs}/(T-t))) \ldots + \text{caseprobs(3,t)}*( \text{Vy}(3,t+1)*(1+1/(T-t-1)) + \ldots \text{cs*EQY(3,t)}/\text{DY}(t)) \ldots + \text{caseprobs(4,t)}*( \text{Vy}(4,t+1) + \ldots \text{cs*EQY(4,t)}/\text{DY}(t) + \text{EQY(4,t)}/\text{DY}(t+1) ); \]

Given the strategy used the value of being at the x patch on average is.

\[ \text{Vavg} = \text{Vexp} * \text{probE}(t) + \text{Vobs} * \text{probO}(t) + \text{Vswt} * \text{probS}(t); \]

\[ \text{Vactions(:,t)} = [\text{Vexp};\text{Vobs};\text{Vswt}]; \]

\[ \text{Vx}(4,t-1) = \text{Vavg}; \]

Which is the value to go of being at the x patch when equilibrium has not been achieved by round t-1.

We have the value of each action, exploit, observe, switch. When the ES is employed all of these will have the same value,
%for now we shift the probabilities of using each action so that actions with high payoff are employed more frequently than actions with low payoff. We do this according to a simple replicator dynamic. The specifics of the dynamic are not important, we are just looking for the fixed point.

newprob0(t) = prob0(t)*(1 - beta*(Vavg-Vobs)/Vavg);
newprobS(t) = probS(t)*(1 - beta*(Vavg-Vswt)/Vavg);
newprobE(t) = 1 - newprob0(t) - newprobS(t);

%This is the average value to go from being at the y patch not at equilibrium on round t-1.
Vy(4,t-1) = ...
caseprobs(1,t)* ( Vx(1,t)*(1+cs/(T-t)) )... 
+ caseprobs(2,t)* ( Vy(2,t)*(1+1/(T-t)) )... 
+ caseprobs(3,t)* ( Vy(3,t+1)*(1+1/(T-t-1)) + ... 
  EQY(3,t)/DY(t) )... 
+ caseprobs(4,t)* ( Vy(4,t+1) + EQY(4,t)/DY(t) + ... 
  EQY(4,t)/DY(t+1));

end;

end;
test = max([(abs(newprob0 - prob0)),max(abs(newprobS - probS))]);
if mod(count,100) == 0 
  %Just to keep an eye on things check test every 100 iterations
  test
end;
probS = newprobS;
prob0 = newprob0;
probE = newprobE;
end;
end
B.5 Program for computing an approximation of $d_{\text{source}}$, $O_{\text{source}}$ and $O_{\text{sink}}$

This first computer program computes an approximation of $D$, the difference in expected total future payoff between a forager at a sink patch of quality $q_h$ and a source patch of quality $q_o$.

```matlab
function gain = gs(qx,qy,D,Obsink,Obsource,M,d,pc,p,co,cs,ps)

%This function computes an approximation of the gain from being at the %y patch instead of the x patch, at equilibrium when the y patch is the %recent sink and the x patch is the recent source.

% $D(qx,qy)$ denotes the equilibrium proportion of foragers at the x patch % given that the x patch was the recent source and the y patch was the % recent sink. Thus a duck should switch to the y patch from the x patch if % $D(qx,qy) < dx$ and should switch to the x patch from the y patch if % $dx < 1-D(qy,qx)$ and should stay where they are if % $1-D(qy,qx) <= dx <= D(qx,qy)$

%We compute this by conditioning on the type of the first change. %The possible types of changes break down into the following cases. % 1 $qy$ gets better and worth switching to $D(qx,nqy) < dx$. % 2 $qx$ gets way better and worth switching $dx < 1-D(qy,nqx)$. % 3 $qx$ gets a little better not worth switching to, % 1-D(qy,nqx) <= dx <= D(nqx,qy) % 4 $qx$ gets worse and worth switching from, $D(nqx,qy) < dx$. % 5 $qy$ gets way worse and worth switching from, $dx < 1-D(nqy,dx)$. % 6 $qy$ gets a little worse but not worth switching from, % 1-D(nqy,dx) <= dx <= D(qx,nqy)

%We cannot compute the the expected value of cases 3 and 6, unless we %condition on the type of the second patch quality change, which we are not %going to do. We can however compute an approximation of the value of %these cases simply by considering how long on average it will be until the %second change.

%ecox is the expected cost of observation factor at the recent source patch %x.
ecox = 1-Obsource(qx,qy)*(1-co);
```
%ecoy is the expected cost of observation factor at the recent sink patch %y.
ecoy = 1-Obsink(qx,qy)*(1-co);
%dx and dy are the equilibrium proportions of ducks at each patch.
dx = D(qx,qy);
dy = 1-dx;

The first contribution to what is gained is that which is accrued before %the first change occurs. The difference in payoff per forager at each %patch at equilibrium is
gain = (qy/dy*ecoy - qx/dx*ecox) * ps/(1-ps);

%pa (probability alive) is the probability that the first change is not a %death, but rather a change in patch quality.
pa = 1 - d/(1-ps);

We then consider the gains made on the round that the first change occurs %scaled by the probability that you are alive then.
gain = gain + pa/2*((qy+1/p)*ecoy/dy-(qx+1/p)*ecox/dx);

% case 1 The quality qy at the sink patch improves and more ducks switch %there.

%probs is the probability with which the ducks at the x patch observe %the y patch.
probs = Obsource(qx,qy);

Some of the ducks at the qx patch will become aware of the change at the %qy patch as soon as it happens.

%Firstobs is the proportion of the total population who are at the x %patch and who observe the y patch quality change right when it %happens.
firstobs = probs*dx;

%secondobs is the proportion of the total population who are at the x %patch and who observe the y patch quality change the round after it %happens, given that all of the ducks who first saw the quality change %have switched.
secondobs = probs*(dx-firstobs);
\% eyb is the expected gain from y getting better, for a duck at the y patch as compared to a duck at the x patch, weighted by the probability that y does get better.

\texttt{eyb = 0;
for i = 1:M-qy
    nqy = qy+i;
}

\%ndx and ndy are the new equilibrium distribution of ducks that will eventually result from this patch quality change.

\texttt{ndy = 1-D(qx,nqy);
}

\%deltady is the proportion of the total population that needs to switch from the x patch to the y patch for the new equilibrium distribution to be achieved.

\texttt{deltady = ndy - dy;
}

\% There were enough observers to bring the system back into equilibrium immediately and those already at the y patch are only ahead the cost of switching there.

\texttt{eyb = eyb + ps * (1-cs)*nqy/ndy * (1-p)^{(nqy-1)*p};
}

\% There were not enough observers to bring the system back into equilibrium right away, but there were enough to bring about equilibrium one round after. The ducks already at the y patch are ahead of the firstobs ducks only by the cost of switching there, but are ahead of all the other ducks at the x patch by the cost of switching there and the benefit from the round out of equilibrium.

\texttt{eyb = eyb + ...
    ((1-firstobs/dx)*...}

\% There were not enough observers to bring the system back into equilibrium on the first two rounds immediately after the
% change, but we assume that there are always enough observers to bring about equilibrium three rounds after a change. This is not always true, but the probability of the cases where it is not true is small enough to ignore here. The ducks already at the y patch are ahead of the first observers only by the cost of switching there, they are ahead of the second observers by the cost of switching there for and the first round spend out of equilibrium, they are ahead of all the other ducks initially at the x patch by those first two rounds out of equilibrium and the cost of switching there.

\[
eyb = eyb + (\ldots (1 - (\text{firstobs} + \text{secondobs})/dx) \times \ldots \\
(\text{ps}^3 * (1 - \text{cs}) * \text{nqy}/\text{ndy} + \ldots \\
\text{ps}^2 * (\text{nqy}/(\text{dy} + \text{firstobs} + \text{secondobs}) - \ldots \\
\text{qx}/(\text{dx} - \text{firstobs} - \text{secondobs})) + \ldots \\
\text{ps} * (\text{nqy}/(\text{dy} + \text{firstobs}) - \text{qx}/(\text{dx} - \text{firstobs}))) + \ldots \\
(\text{secondobs}/dx) \times \ldots \\
(\text{ps}^2 * (1 - \text{cs}) * \text{nqy}/(\text{dy} + \text{firstobs} + \text{secondobs}) + \ldots \\
\text{ps} * (\text{nqy}/(\text{dy} + \text{firstobs}) - \text{co*qx}/(\text{dx} - \text{firstobs}))) + \ldots \\
(\text{firstobs}/dx) \times \ldots \\
(\text{ps} * (1 - \text{cs}) * \text{nqy}/(\text{dy} + \text{firstobs}))) \times \ldots \\
(1 - p) \times (\text{nqy} - 1) * p; \\
\text{end;}) \text{end;} \\
\text{end;}
\]

gain = gain + pa/2*eyb;

% case 2 qx gets way better and worth switching to, dx < 1-D(qy,nqx). The ducks who are already at the x patch will have the same expected payoff as the ducks that join them less the cost of switching there.

%We first need to find the critical value, bqx for which the x patch has improved past in order for the x patch to flip from being the source patch to being the sink patch.

bqx = qx;
while dx > (1-D(qy,bqx))
    bqx = bqx + 1;
    if bqx > M 
        break
end;
end;
if bqx > qx
    bqx = bqx - 1;
end;

% exb is the expected loss from y getting better, for a duck at the y patch
% as compared to a duck at the x patch, weighted by the probability that x
% does get better enough to be worth switching to.

%probobs is the probability with which the ducks at the y patch observe
%the x patch.
probobs = Obsink(qx,qy);

%firstobs is the proportion of the total population who are at the y
%patch and who observe the x patch quality change right when it
%happens.
firstobs = probobs*dy;

%secondobs is the proportion of the total population who are at the y
%patch and who observe the x patch quality change the round after it
%happens, given that all of the ducks who first saw the quality change
%have switched.
secondobs = probobs*(dy-firstobs);

exb = 0;
for i = 1:M-bqx
    nqx = bqx+i;
    
    %ndx and ndy are the new equilibrium distribution of ducks that
    %will eventually result from this patch quality change.
    ndx = 1-D(qy,bqx);

    %deltadx is the proportion of the total population that needs to
    %switch from the x patch to the y patch for the new equilibrium
    %distribution to be achieved.
    deltadx = ndx - dx;

    if firstobs > deltadx
APPENDIX B. COMPUTER PROGRAMS

\% There were enough observers to bring the system back into
\% equilibrium immediately and those already at the x patch are
\% only ahead the cost of switching there.
\texttt{exb} = \texttt{exb} + \texttt{ps} * (\texttt{1-Cs}) * \texttt{nqx}/\texttt{ndx} * (\texttt{1-p})^\texttt{(nqx-1)} * \texttt{p};

\textbf{elseif} \ \texttt{firstobs} + \texttt{secondobs} > \texttt{deltadx}
\%
% There were not enough observers to bring the system back into
% equilibrium right away, but there were enough to bring about
% equilibrium on round after. The ducks already at the x patch
% are ahead of the \texttt{firstobs} ducks only by the cost of switching
% there, but are ahead of all the other ducks at the y patch by
% the cost of switching there and the benefit from the round
% out of equilibrium.
\texttt{exb} = \texttt{exb} + ...
\texttt{(1-firstobs)/dy)*...}
\texttt{(ps^2 * (1-Cs)*nqx/ndx + ...
ps * (nqx/(dx+firstobs) - qy/(dy-firstobs))) + ...
(firstobs/dy)*...}
\texttt{(ps * (1-Cs)*nqx/(dx+firstobs))) * ...}
\texttt{(1-p)^/(nqx-1)*p;}

\textbf{else}
\%
% There were not enough observers to bring the system back into
% equilibrium on the first two rounds immediately after the
% change, but we assume that there are always enough observers
% to bring about equilibrium three rounds after a change. This
% is not always true, but the probability of the cases where it
% is not true is small enough to ignore here. The ducks already
% at the x patch are ahead of the first observers only by the
% cost of switching there, they are ahead of the second
% observers by the cost of switching there for and by the first
% round spend out of equilibrium, they are ahead of all the
% other ducks initially at the y patch by those first two
% rounds out of equilibrium and the cost of switching there.
\texttt{exb} = \texttt{exb} + (...}
\texttt{(1 - (firstobs+secondobs)/dy)*...}
\texttt{(ps^3 * (1-Cs)*nqx/ndx + ...}
\texttt{ps^2 * (nqx/(dx+firstobs+secondobs) - ...}
\texttt{qy/(dy-firstobs-secondobs)) + ...}
\texttt{ps * (nqx/(dx+firstobs) - qy/(dy-firstobs))) + ...}
\texttt{(secondobs/dy)*...}
\texttt{(ps^2 * (1-Cs)*nqx/(dx+firstobs+secondobs) + ...}
APPENDIX B. COMPUTER PROGRAMS

\[ ps \ast (nqx/(dx+firstobs) - co*qy/(dy-firstobs))) + \ldots \]
\[ (firstobs/dx)\ast\ldots \]
\[ (ps \ast (1-cs)*nqx/(dx+firstobs))) \ast\ldots \]
\[ (1-p)^(nqx-1)*p; \]
end;
end;

gain = gain - pa/2*exb;

%case 3 qx gets a little better but not worth switching to, that is qx <=
%nqx <= bqx. We compute the expected gain from being at the y patch instead
%of the x patch when a change like this occurs, weighted by the probability
%of such a change occurring.

%nqx satisfies qx <= nqx <= bqx with probability.
probcase3 = (1-p)^(qx-1) - (1-p)^bqx;

%the expected value of nqx in this case is
n = bqx - (qx-1);
index = 1:n;
enqx = (qx-1) + p/(1-(1-p)^n)*(1-p).^(index-1)*index';
exlb = (enqx/dx*ecox - qy/dy*ecoy)*ps/(1-ps)*probcase3;
gain = gain - pa/2*exlb;

%case 4 qx gets worse and worth switching from, nqx < qx.

exw = 0;
for i = 1:qx-1
    nqx = i;

    %ndx and ndy are the new equilibrium distribution of ducks that will
    %result from this patch quality change.
    ndy = 1-D(nqx,qy);

    %Unlike the case when a patch improves and it is up to the ducks who
%are not there to learn about this change and react to it, when a patch
%gets worse all of the ducks there know about it immediately, and since
%no ducks have joined them from the other patch they know that the
%other patch has the same value it did last time they checked in on it
%(or close to) and so we assume that equilibrium is achieved
APPENDIX B. COMPUTER PROGRAMS

% immediately.
exw = exw + ps * (1-cs)*qy/ndy * (1-p)^(nqx-1)*p;
end;
gain = gain + pa/2*exw;

% case 5 qy gets way worse and worth switching from, dx < 1-D(nqy,dx). We
% compute the expected loss from being at the y patch instead of the x patch
% in this situation, weighted by the probability of this case.

% We first need to compute the critical value wqy such that if nqy is as bad
% or worse than wqy, then it is worth leaving the y patch for the x patch.

wqy = qy;
while (1-D(wqy,qx)) < dx
    wqy = wqy - 1;
    if wqy == 0
        break
    end;
end;

eyw = 0;
for i = 1:wqy
    nqy = i;

    %ndx and ndy are the new equilibrium distribution of ducks that will
    % result from this patch quality change.
    ndx = 1-D(nqy,qx);

    % Unlike the case when a patch improves and it is up to the ducks who
    % are not there to learn about this change and react to it, when a patch
    % gets worse all of the ducks there know about it immediately, and since
    % no ducks have joined them from the other patch they know that the
    % other patch has the same value it did last time they check in on it
    % (or close to) and so we assume that equilibrium is achieved
    % immediately.
    eyw = eyw + ps * (1-cs)*qx/ndx * (1-p)^(nqy-1)*p;
end;
gain = gain - pa/2*eyw;
APPENDIX B. COMPUTER PROGRAMS

%%% case 6 qy gets a little worse but not worth switching from, that is
%%% wqy <= nqy <= qy. As in case three we compute the expected gain from
%%% being at the the y patch instead of the x patch when this kind of change
%%% happens.

%%% nqy satisfies wqy < nqy <= qy with probability.
probcase6 = (1-p)^wqy - (1-p)^qy;

%%% the expected value of nqx in this case is
n = qy - wqy;
index = 1:n;
enqy = wqy + p/(1-(1-p)^n) * (1-p).^(index-1) * index';
eylw = (enqy/dy*ecoy - qx/dx*ecox)*ps/(1-ps)*probcase6;
gain = gain + pa/2*eylw;
end

These next two computer programs compute an approximation of the difference
in expected total future payoff between a forager that observes at a sink patch and a
forager that exploits at a sink patch, and the differen in expected total future payoff
between a forager that observes at a source patch and a forager that exploits at a
source patch.