The Border Patrol Game

Matthew Darlington, B.Sc.(Hons.), M.Res

Submitted for the degree of Doctor of

Philosophy at Lancaster University.

24th April, 2024

STOR-i

excellence with impact

Abstract

The issue of how best to patrol a border can be found in many settings. Important examples include protecting important infrastructure such as airports, preventing the smuggling of illicit items, or defending computers in a network. This thesis contributes to the existing literature by developing two unique models for border patrol.

We begin by introducing a model where a group of smugglers play a game against a single patroller. We investigate how communication and cooperation between the smugglers affect the equilibria in the game. Smugglers are located at different locations along a border and, for each smuggler, a decision is made about whether they will attack. Simultaneously, the patroller chooses one of the locations to defend. Smugglers obtain rewards for making successful attacks, but incur penalty costs if they are caught by the patroller. The reward to an individual smuggler for making a successful attack decreases with the total number of successful attacks made, so that the smugglers obtain diminishing marginal returns as they smuggle larger quantities of items. We define equilibria in three different cases: selfish smugglers without communication, selfish smugglers with communication, and cooperative smugglers. We study the equilibria in each case and establish properties of the associated smuggler and patroller strategies. We show that communication and cooperation both tend to improve the smugglers' expected returns, while (perhaps counter-intuitively) the smugglers attack less often when they are cooperating than when they are communicating but acting selfishly.

Our second model considers a similar problem to the first, but we make some im-

portant changes to the game. We simplify the payoff structure of the model, however, we are then able to study a repeated game with inter-temporal dependence. The application of the model is examining a group of cooperating smugglers who make regular attempts to bring small amounts of illicit goods across a border. A single patroller has the goal of preventing the smugglers from doing so, but must pay a cost to travel from one location to another. We model the problem as a two-player stochastic game and look to find the Nash equilibrium to gain insight to real world problems. Our framework extends the literature by assuming that the smugglers choose a continuous quantity of contraband, complicating the analysis of the game. We discuss a number of properties of Nash equilibria, including the aggregation of smugglers, the discount factors of the players, and the equivalence to a zero-sum game. Additionally, we present algorithms to find Nash equilibria that are more computationally efficient than existing methods. We also consider certain assumptions on the parameters of the model that give interesting equilibrium strategies for the players.

Furthermore, we introduce a multiple patroller extension to the second model. The addition of multiple patrollers increases the complexity of the game, and exactly finding Nash equilibria becomes an even more complex task. We describe how we model the multiple patroller extension, and detail why previous methods for the single patroller game no longer apply. Three different techniques to study the game are then discussed. Firstly, we look at a method using subgradient descent to try to find the exact Nash equilibrium in the game. Secondly, we look at two different heuristics for the patroller's strategy. We first consider the myopic strategy for the patrollers, and then introduce a method of partitioning the border into multiple segments each defended by one patroller. The performance of the heuristics is numerically investigated and used to evaluate the convergence of the subgradient method. Finally, we discuss how reinforcement learning can be applied to our model. We consider two different reinforcement learning approaches, fictitious play and Q-learning, and then provide a numerical analysis of the

resulting player behaviours. We conclude with several directions of further work for the multiple patroller problem.

Acknowledgements

Firstly, I would like to thank my supervisors David, Kevin, Rob and Roberto for all their help over the past three and a half years. I've learnt so much through the process and it has been a pleasure to work with you all.

The STOR-i CDT has been an amazing place to spend the last five years, and I would like to thank Jon, Idris, Kevin, Nicky, Kim, and Wendy for their work running the centre. From my internship to my master's degree and then the PhD there has always been so much support from the staff that has allowed me to just focus on my work. I would also like to thank everyone at the Naval Postgraduate School that helped make my three visits there possible. My time in Monterey has been an incredible highlight from my PhD.

I've had a wonderful time at STOR-i and I would like to thank everyone within the centre making it so enjoyable. I'd like to especially thank Ed, Hamish, Peter, Matt Randle and Tamás for all the great times we've had since we joined.

I would like to thank all my family for all their support during my many years of school and university. Finally, I would like to thank Alicia for her support through my PhD, from encouraging me to sign up at the start to helping proofread at the end.

Declaration

I declare that the work in this thesis has been done by myself and has not been submitted elsewhere for the award of any other degree.

A version of Chapter 4 has been submitted for publication by Darlington, M., Glazebrook, K. D., Leslie, D. S., Shone, R., and Szechtman, R. in 2024.

A version of Chapter 5 has been published as Darlington, M., Glazebrook, K. D., Leslie, D. S., Shone, R., and Szechtman, R. (2023). A stochastic game framework for patrolling a border. European Journal of Operational Research, 311(3):1146–1158.

Matthew Darlington

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

There are a number of examples from around the world where a border needs to be protected from an adversary. Deciding how to best utilize the constrained resources to defend the border is a paramount issue that affects government organizations worldwide. A selection of key issues which are covered by border patrolling include:

- Drug trafficking through Europe (Baniya, 2023)
- Oil smuggling (Savage and Bergman, 2023)
- Illicit trade of wildlife (Freedman, 2022)
- Drug trafficking across the U.S. - Mexico border (Gutierrez and Henkel, 2021)
- Illegal fishing in the continental shelf off South America (Goodman, 2021)

Designing an effective patrol pattern is a difficult challenge, which needs to take into account many considerations. Firstly, due to constraints on resources, the patroller cannot defend everywhere at once. Therefore, there needs to be an intelligent decision maker choosing which areas of the border will be patrolled, and consequently, which

will be left exposed. Secondly, it is intuitive that there should also be some source of randomness within the decision. If the patroller repeatedly made the same deterministic decisions, then the smugglers could observe and learn where to attack. Furthermore, there is even the possibility for exploiting inside knowledge if the deterministic plans were leaked. Having stochastic patrol patterns builds resilience against such factors by making it harder for the smugglers to predict their actions.

There has been a wide range of border patrol problems tackled in the literature. Due to specific problems having a unique set of challenges to consider, there are a number of distinct models for particular applications. However, the majority of models use a game theoretic framework with which to capture the decision-making process. Game theory is the study of how multiple rational players make actions to best meet their own individual objectives. In border patrol problems, there is at least a defending side and an attacking side, with further possibilities for more complex dynamics.

In this thesis, we consider a scenario where patrollers attempt to stop a group of smugglers taking items across a border. We think of a border as being represented by a finite set of locations which could correspond to roads, border control posts or a discretized section of air, land, or sea. The smugglers aim to send some illicit items through these locations, whilst the patroller attempts to catch or deter them. It is assumed that either side receives some known fixed rewards and penalties depending on the quantities sent, captured, and smuggled. Given the described framework, we study three questions about the behavior of the patroller and the smugglers within the border patrol game.

The first question considers how communication and cooperation between smugglers affects equilibria in the game. It is a common assumption that if there are multiple adversaries, that they will be working together against the patroller. The assumption has the benefit of being the worst-case scenario for the patroller. However, we show that by ignoring other possibilities, we are excluding the analysis of interesting potential

behavior from the smugglers. The thesis introduces a modelling framework for these settings, and details the equilibria for each case. By considering this problem, the contributions to the literature include:

- A new model is given for a border patrol problem, which considers how smugglers receive diminishing returns for trafficking increasing amounts of illicit items.
- We consider three cases of smuggler behavior. Firstly, selfish smugglers that work on their own to maximize their own reward. Secondly, selfish smugglers with communication which, while still only maximizing their own reward, can broadcast their actions. Finally, we look at cooperative smugglers that work together to maximize the total reward across the whole group.
- It is shown that at equilibrium, across all cases of smuggler behavior, the smugglers need to attempt to traffic items with equal probability. Additionally, in two of the three cases, we prove that at equilibrium the patroller must defend every location with equal probability.
- In two of the three cases of smuggler behavior, we prove analytically which strategies are equilibria. Furthermore, we detail all possible equilibria in the remaining case, subject to additional assumptions on the patroller's behavior.
- We show that allowing the smugglers to cooperate has a perhaps unanticipated consequence of decreasing the number of attacks made. Furthermore, we can also sometimes see this behavior when the smugglers gain the ability to communicate.
- Finally, we discuss how the penalty given to caught smugglers affects the equilibria of the game under different settings of smuggler behavior.

The second question considers what the optimal patrolling strategies are if there are movement costs for the patroller, and the game takes place over multiple time steps.

There is a trade-off required between protecting the border in the immediate decision and planning for future actions. The thesis introduces a stochastic game framework for this problem and proves how to compute the equilibria under certain conditions.

The contributions to the literature achieved by studying this problem include:

- Our model introduces a continuous set of actions available to the smugglers, whilst similar models in the literature only allow for a finite set of possibilities. In reality, smugglers have an extensive number of options available to them, and so our continuous action space allows for a more realistic representation of the problem.
- We analytically prove properties about the Nash equilibria of the game. The results provide insight into the behavior of how rational players act in our model, and therefore how they might act in a real-life patrolling scenario.
- New algorithms are developed to overcome the computational challenges of finding Nash equilibria in our model. We show that our algorithms find the optimal solution, and furthermore, they do so in a shorter time than existing methods where comparisons can be made.

Our third question is to examine what happens when we extend the framework of the second question by allowing the patrolling side to have more than one patroller. Multiple patrollers is a realistic assumption when government organizations will have large teams to catch smugglers. This thesis considers methods to exactly find equilibria, heuristics to approximate them and also reinforcement learning approaches to the problem. The contributions presented include:

- We introduce a new stochastic game framework for patrolling a border, which allows for multiple patrollers to cooperate in defending.
- Algorithms that find Nash equilibria are discussed, implemented and tested on examples. However, we also show these methods become too computationally expensive to use on larger problems.

- We consider two intuitive heuristics, myopic patrollers and partitioning the border, and compare their performance against our exact methods where possible.

1.2 Thesis Outline

The remainder of the thesis consists of seven chapters. The opening two chapters contain a literature review, the middle three chapters provide contributions to the literature, and the final chapter concludes the thesis.

Chapter 2 provides a literature review for the area of game theory. Both normal-form and stochastic games are discussed, including definitions for equilibria and how to compute them. To conclude, an overview of reinforcement learning in games is presented.

Chapter 3 considers the literature of border patrol problems that have previously been studied. There are three main sections in the chapter, each looking at a different model that has both had a significant impact in its area and is close to the problems considered in this thesis. A wider overview of the literature is then presented, giving a broader look at the range of problems previously examined.

Chapter 4 gives a study of cooperation and communication between smugglers in our border patrol game. First, we define how concepts of equilibria look depending on how the smugglers work together. Then, key insights are given showing how these assumptions affect the behavior at equilibrium between the players in the game.

Chapter 5 considers a stochastic game framework for patrolling borders, the important modification being a movement cost penalty applied to the patroller. We first define how the game will be played, give insights to the equilibria, and moreover, detail how equilibria can be efficiently computed.

Chapter 6 looks to extend the work of Chapter 5 by adding in multiple patrollers to the game. We look at how the equilibria can be exactly found, introduce heuristics

to ease the computational effort, and furthermore, look at how learning dynamics can converge to the equilibrium.

Finally, Chapter 7 concludes the thesis providing a brief summary of the contributions made, along with a number of directions for further work on the topic.

Chapter 2

Games

Game theory allows us to study interactions between multiple decision makers, or players, where their decisions impact the others. The objective of a player cannot simply be to maximize their own payoff, since they must take into consideration the actions of the other players. This chapter contains a review of both normal-form games and their extension into stochastic games. Finally, an introduction to reinforcement learning in games in presented.

2.1 Normal-form Games

Normal-form games represent a one-off decision-making problem, consisting of three elements. Firstly, there is a set of n players involved in the game, denoted by player i for $i \in [n]$. Secondly, each player has a set of actions they can choose between, denoted by $\mathcal{A}_i$ for player i. Finally, each player has a payoff from the game which depends on the actions chosen by every player. Given that the players choose actions $\boldsymbol{a} = (a_1, \ldots, a_n)$, we denote player i's payoff as $r_i(\boldsymbol{a})$.

The number of players in the game could be a finite integer or infinite, but it is common to study two player games. Two player games are more easily analyzed, since it reduces the number of opponents which need to be considered. In a two player

normal-form game, we can represent the payoffs to the players as matrices M_1 and M_2, where the $(i,j)th$ entry is given by $r_1(a_i, a_j)$ and $r_2(a_i, a_j)$ respectively. The remainder of this section will focus on the study of two player normal-form games, unless specified.

An important subset of two player normal-form games are those where the payoff to one player is equal to the negative of the payoff to the other player. We therefore have for every action $a_i \in \mathcal{A}_1$ and $a_j \in \mathcal{A}_2$ that $r_1(a_i, a_j) = -r_2(a_i, a_j)$. These games are called zero-sum games.

2.1.1 Nash Equilibria

v. Neumann (1928) first introduced the concepts of equilibria in games. It was later that Nash (1950) generalized the notion of equilibria to all normal form games, now called Nash equilibria. The notion of a Nash equilibrium is that given all the strategies are known, no player has an incentive to deviate away from their strategy. A player would only have incentive to deviate from their strategy if they can play a different action which would strictly increase their expected reward.

We begin by defining what a strategy in a normal-form game is. A strategy for a player i is a probability distribution over their action set $\mathcal{A}_i$, denoted by π_i. The probability that a player i chooses an action $a_i \in \mathcal{A}_i$ can then be given by $\pi_i^{a_i}$. We now define a Nash equilibrium in definition 2.1.1.

Definition 2.1.1 (Nash Equilibrium)**.** *Consider a finite n player normal-form game. The strategies $\pi_1, \ldots, \pi_n$ are a Nash equilibrium if no player has incentive to deviate from their strategy given the other players' strategy is fixed. A player has incentive to deviate if they can strictly increase their expected reward by changing strategy and so for every player i,*

$$\sum_{a_1 \in \mathcal{A}_1} \pi_1^{a_1} \cdots \sum_{a_i \in \mathcal{A}_i} \pi_i^{a_i} \cdots \sum_{a_n \in \mathcal{A}_n} \pi_n^{a_n} r_i(\boldsymbol{a}) \geq \sum_{a_1 \in \mathcal{A}_1} \pi_1^{a_1} \cdots \sum_{a_i \in \mathcal{A}_i} \tilde{\pi}_i^{a_i} \cdots \sum_{a_n \in \mathcal{A}_n} \pi_n^{a_n} r_i(\boldsymbol{a})$$

for all $\tilde{\pi}_i \in \Delta(\mathcal{A}_i)$.

Note that if the two players choose strategies π_1 and π_2 respectively, then the expected payoff to a player i can be written as $(\pi_1)^T M_i \pi_2$. We can therefore simplify the definition of a Nash equilibrium in the two player setting.

Definition 2.1.2 (Nash Equilibrium)**.** *Consider a finite two player normal-form game with payoff matrices* M^1 *and* M^2*. The strategies* π_1 *and* π_2 *are a Nash equilibrium if no player has incentive to deviate from their strategy given the other player's strategy is fixed. A player has incentive to deviate if they can strictly increase their expected reward by changing strategy, and so,*

$$(\boldsymbol{\pi}_1)^T M_1 \boldsymbol{\pi}_2 \geq (\tilde{\boldsymbol{\pi}}_1)^T M_1 \boldsymbol{\pi}_2 \text{ for all } \tilde{\boldsymbol{\pi}}_1 \in \Delta(\mathcal{A}_1)$$
$$(\boldsymbol{\pi}_1)^T M_2 \boldsymbol{\pi}_2 \geq (\boldsymbol{\pi}_1)^T M_2 \tilde{\boldsymbol{\pi}}_2 \text{ for all } \tilde{\boldsymbol{\pi}}_2 \in \Delta(\mathcal{A}_2)$$

2.1.2 Calculating Nash Equilibria

Having defined Nash equilibria in normal-form games, we now look to how we can compute them. In general, the analysis of Nash equilibria in a normal-form game is a difficult challenge. There may not exist any Nash equilibria in the game, or there may be an infinite number of them. However, in the case of two player zero-sum normal form games, there always exists a Nash equilibrium due to the result of v. Neumann (1928).

Theorem 2.1.3 (von Neumann Minimax Theorem)**.** *Every finite, two player, zero-sum normal-form game has a Nash equilibrium. The strategies* π_1 *and* π_2 *that form a Nash equilibrium are the solution to the following 'minimax' optimization problem,*

$$\max_{\pi_1} \min_{\pi_2} \ (\pi_1)^T M_1 \pi_2 = \min_{\pi_2} \max_{\pi_1} \ (\pi_1)^T M_1 \pi_2$$

It was further proven that it is not necessary for player 2 to minimize over their

space of strategies, but instead it is sufficient to minimize over their actions. The result is shown in Lemma 2.1.4, and presented as found in Filar and Vrieze (2012).

Lemma 2.1.4. *In a finite, two player, zero-sum normal form game if* $\boldsymbol{\pi}^1$ *and* $\boldsymbol{\pi}^2$ *we have,*

$$\max_{\pi_1} \min_{\pi_2} (\pi_1)^T M_1 \pi_2 = \max_{\pi_1} \min_{j} (\pi_1)^T A \boldsymbol{e}_j \tag{2.1.1}$$

where e_j *is a vector of zeroes with a one in the* j*th entry.*

We denote the expected value of the game in a Nash equilibrium to the two players as V_1 and, V_2 respectively. Note, since the game is zero-sum, we have that $V_1 = -V_2$. We can formulate the minimax optimization problem found in (2.1.1) by using a linear program, found in (2.1.2). Linear programs are optimization problems with a linear objective function and linear constraints. There is a wide literature focussed on finding efficient algorithms to solve linear programs, an overview of which can be found in Chvátal (1983).

$$\begin{aligned} \text{maximise} \quad & V_1 \\ \text{such that} \quad & V_1 \leq (\pi_1)^T A \boldsymbol{e}_j \; \forall \; j \\ & \pi_1^{a_i} \in [0,1] \\ & \sum_{a_i \in \mathcal{A}_\rangle} \pi_1^{a_i} = 1 \end{aligned} \tag{2.1.2}$$

The objective of the linear program is to pick a strategy for player 1 that maximizes the value of the game to them. However, the first constraint ensures that for any action a_j picked by player 2 the value of the game does not exceed the expected payoff. The second and third constraints ensure that player 1 picks a valid strategy.

2.1.3 Correlated Equilibria

Nash equilibria are not the only type of equilibrium that can be found in a normal-form game. A different kind of equilibria are correlated equilibria, first introduced by Aumann (1987). A correlated equilibrium consists of some publicly available stochastic signal which is displayed to each player in the game. Every player then maps this public signal to a private action, which they play in the game. A signal and a set of mappings for each player form a correlated equilibrium if for any player, given all other mappings remain fixed, they do not have any incentive to deviate from their mapping.

Every Nash equilibrium gives a correlated equilibrium. However, the converse is not true and there can exist correlated equilibria which do not correspond to any Nash equilibrium. To find a correlated equilibrium, given that there exists a Nash equilibrium, we can construct a signal by taking the strategies of the Nash equilibrium, and give each player the identity mapping from the public signal to their action. We have a correlated equilibrium, since assuming a Nash equilibrium guarantees that no player has any incentive to deviate away from their strategy.

2.2 Stochastic Games

Stochastic games are a generalization of normal-form games that allow for repeated actions to be taken whilst a probabilistic transition alters the state of the game. In this thesis, we consider discounted stochastic games which consist of six elements. Firstly, a set of n players are involved in the game, denoted by player i for $i \in [n]$. Secondly, a state space $\mathcal{S}$ which represents all possible states the game can be in. Thirdly, each player has a set of actions they can choose, which possibly depends on the state of the game s, denoted by $\mathcal{A}_i(s)$ for player i. Each player has a payoff from the game which depends on the actions chosen by every player, and the state of the game s. Given that the players choose actions $\boldsymbol{a} = (a_1, \ldots, a_n)$, we denote player i's payoff as $r_i(s, \boldsymbol{a})$.

Furthermore, there is a probability distribution for every combination of state s and actions $\boldsymbol{a}$ which determines the transition for the next step given by P. To determine the initial starting state of the game, a probability distribution P_0 is specified. Finally, there is a common discount rate, $\gamma \in [0, 1)$ which represents the trade-off for the players between immediate and future rewards.

As in the normal-form game setting, whilst we could have any number of players in the stochastic game, we will primarily restrict our focus to two player stochastic games. Additionally, the zero-sum two player stochastic games where the payoff function of one player is equal to the negative of the payoff function of the other player. Thus, in two player zero-sum stochastic games we have $r_1(s, a_1, a_2) = -r_2(s, a_1, a_2)$ for all states $s \in \mathcal{S}$ and actions $a_1 \in \mathcal{A}_1$, $a_2 \in \mathcal{A}_2$.

2.2.1 Nash Equilibria

We now look to define Nash equilibria for stochastic games. Previously, in the normal-form game setting, the players chose strategies which were probability distributions over their actions. In the stochastic game setting, this notion is extended and strategies are now a probability distribution over the players actions but depend on the current state of the game. Thus, a strategy for a player i is denoted by $\Pi_i = (\pi_i(s))_{s \in \mathcal{S}}$.

Having defined the players' strategies, we now need to define the expected reward to each player. Recall, in the normal-form game it can be easily calculated by taking the expectation of the payoff with respect to the player's strategies. The same definition is used in the stochastic game setting, but we take the expectation over the entire infinite discounted horizon. We denote the steps of time in which players choose actions by $t = 1, 2, \ldots$. The expected reward to a player i given the two players take strategies Π_1 and Π_2 is given by,

$$U_i(\Pi_1, \Pi_2) = \mathbb{E}_{\Pi_1, \Pi_2, P_0, P}\left[\sum_{t=0}^{\infty} \gamma_i^t r_i(s_t, a_1, a_2)\right]$$

The expectation is taken over the random strategies of both players, Π_1 and Π_2, the probability distribution of the first state, P_0, and the stochastic transitions between states given by P.

We can now define a Nash equilibrium for a two player stochastic game.

Definition 2.2.1 (Nash Equilibrium)**.** *Consider a finite two player stochastic game. The strategies* Π_1 *and* Π_2 *are a Nash equilibrium if no player has an incentive to deviate from their strategy given the other player's strategy is fixed. A player has incentive to deviate if they can strictly increase their expected reward, and so,*

$$U_1(\Pi_1, \Pi_2) \geq U_1(\tilde{\Pi}_1, \Pi_2) \; \forall \; \tilde{\Pi}_1$$
$$U_2(\Pi_1, \Pi_2) \geq U_2(\Pi_1, \tilde{\Pi}_2) \; \forall \; \tilde{\Pi}_2$$

Note that a Nash equilibrium as defined in Definition 2.2.1 is a Nash equilibrium with respect to any initial distribution over the starting state of the game.

In general, it is a difficult problem to find Nash equilibria in two player stochastic games. Therefore, we restrict our focus to two player, zero-sum stochastic games since more can be analyzed in this setting.

Shapley (1953) proved that in finite, two player, zero-sum stochastic games, a Nash equilibrium always exists. Given a Nash equilibrium with strategies Π_1 and Π_2, we define the value of a state s to a player i as the expected reward given the game begins in that state. We denote the value of the state s to a player i as $V_i(s)$ where,

$$V_i(s) = U_i(\Pi_1, \Pi_2)\Big|_{s_0=s} = \mathbb{E}_{\Pi_1,\Pi_2}\left[\sum_{t=0}^{\infty} \gamma_i^t r_i(s_t, a_1, a_2) \;\middle|\; s_0 = s\right]$$

Since we are only considering two player, zero-sum stochastic games, we drop the dependence on the player for ease of notation, since $V_1(s) = -V_2(s)$. From this point, unless specified otherwise, we will be referring to player one. Shapley proved in his 1953 paper that the value of a finite, two player, zero-sum stochastic game is the

unique solution to the system of equations,

$$V(s) = val\left[r(s, a_1, a_2) + \gamma \sum_{s' \in \mathcal{S}} V(s')\mathbb{P}(s' \mid s, a_1, a_2)\right]$$
$$= \max_{\pi_1} \min_{\pi_2} \left\{ \sum_{a_1 \in \mathcal{A}_1(s)} \pi_1^{a_1} \sum_{a_2 \in \mathcal{A}_2(s)} \pi_2^{a_2} \left[r(s, a_1, a_2) + \gamma \sum_{s' \in \mathcal{S}} V(s')\mathbb{P}(s' \mid s, a_1, a_2)\right] \right\}$$

where $val[A]$ is an operator denoting the value of the two player, zero-sum normal form game with payoff matrix A. Note, this is similar to the Bellman equations for a Markov decision process, but we take the minimax of both players' actions rather than maximizing over one player's actions.

Shapley (1953) proved that we can calculate the value of each state by iteratively solving the system of equations, given any choice of starting values. Starting from arbitrary state values $V^0(s)$, we define $V^i(s)$ to be

$$V^i(s) = val\left[r(s, a_1, a_2) + \gamma \sum_{s' \in \mathcal{S}} V^{i-1}(s')\mathbb{P}(s' \mid s, a_1, a_2)\right].$$

As the operator giving the value of the game is a contraction mapping, we have that as $i \to \infty$ we know $V^i(s) \to V(s)$ for every state s. Since we previously saw the value of a normal-form two player zero-sum game can be calculated with a linear program, we can now compute the values of each state in a two player, zero-sum stochastic game. We can repeatedly solve for the values of each state, until we meet some user chosen convergence target. A common method to check for convergence is to pick a small threshold $\epsilon > 0$ stop when

$$\max_s \|V^i(s) - V^{i-1}(s)\| < \epsilon.$$

2.2.2 Single Controller Stochastic Games

We now introduce a class of stochastic games called single controller stochastic games. The single controller property is when only a single player can influence the state transitions in the stochastic game. Therefore, in the two player case if

$$\mathbb{P}(s' \mid s, a_1, a_2) = \mathbb{P}(s' \mid s, a_1, \tilde{a}_2)$$

for any $s, s' \in \mathcal{S}$, $a_1 \in \mathcal{A}_1(s)$ and $a_2, \tilde{a}_2 \in \mathcal{A}_2(s)$ we have that player one is the single controller of the game.

There are algorithms in the literature which can exactly compute the Nash equilibria in zero-sum, single controller stochastic games, an example being found in Raghavan (2003). If the zero-sum stochastic game has player one as the single controller, there exists a linear program to find the state values and player two's strategy in a Nash equilibrium.

$$\begin{aligned}
\text{maximise} \quad & \sum_{s \in \mathcal{S}} V(s) \\
\text{such that} \quad & V(s) \geq \sum_{a_2 \in \mathcal{A}_2(s)} r_2(s, a_1, a_2) \pi_2^{a_2}(s) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}(s' \mid s, a^1) V(s') \; \forall \; s \in \mathcal{S}, a^1 \in \mathcal{A}_1(s) \\
& \pi_2^{a_2}(s) \in [0, 1] \; \forall \; s \in \mathcal{S}, a_2 \in \mathcal{A}_2(s) \\
& \sum_{a_2 \in \mathcal{A}_\in} \pi_2^{a_2}(s) = 1 \; \forall \; s \in \mathcal{S}
\end{aligned}$$

The objective of the linear program maximizes the value of the game to player two. The first constraint ensures that whatever player one's action is, they cannot do better than the value of the game in the objective function. The final two constraints ensure that player two's strategy is properly defined.

2.3 Reinforcement Learning

Reinforcement learning allows for an intelligent agent to learn from their environment to take decisions. Typically, in reinforcement learning a single agent is considered, with the environment being modelled as a Markov decision process. An overview of reinforcement learning in Markov decision processes can be found in Sutton and Barto (2018). However, we would like to instead consider the application of reinforcement learning to games with multiple decision makers.

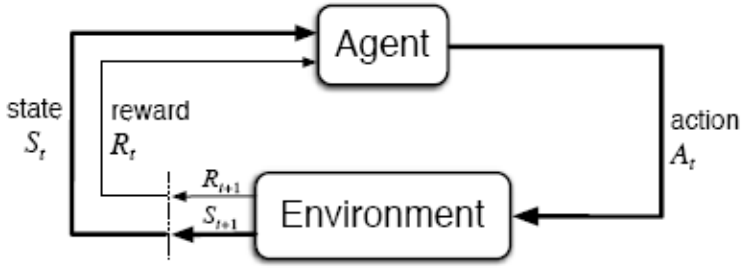


Figure 2.3.1: The agent-environment interaction Sutton and Barto (2018)

The classic framework of a reinforcement learning algorithm is displayed in Figure 2.3.1. Each agent first observes the state of its environment, and then chooses an action. The actions are played, resulting in both a reward being given to each agent and the environment changing state. Information is then given back to each agent, who can then learn from their experience. Learning from experience in this context refers to the updating of some parameters, using a set of rules given to the agents.

In this section, we will first discuss reinforcement learning in normal-form games, before moving onto considering stochastic games.

2.3.1 Normal-form Games

An example of a reinforcement learning algorithm in a normal-form game is fictitious play. Fictitious play was first introduced by Brown (1951) and is a learning rule that assumes every opponent is playing their actions from some stationary strategy. There-

fore, as the actions of the opponents are observed, each player can update their belief of the opponent's fixed strategy. However, if multiple players are learning in such a manner then the assumption of a stationary strategy is clearly unsatisfied, which can cause issues of convergence.

Fictitious play has been proven to converge to a Nash equilibrium in two player games under a number of assumptions including zero-sum finite games by Robinson (1951), potential games by Monderer and Shapley (1996), and non-zero sum games where one player has at most two actions Berger (2005). However, there are simple examples (such as a generalized rock-paper-scissors-game by Shapley (1964)) that do not converge to a Nash equilibrium, and instead the strategies have a cyclic behavior.

2.3.2 Stochastic Games

Our main interest in reinforcement learning is its application to two player zero-sum stochastic games. In later sections of the thesis, we introduce stochastic games where computing the Nash equilibrium exactly becomes too computationally intensive. However, we can instead put agents in the environment to learn Nash equilibrium with reinforcement learning. We present two reinforcement learning algorithms that have recently been published in the literature, the former being a fictitious play algorithm and the latter based on Q-learning.

Firstly, we introduce the notion of a Q-function. Suppose that player i knows that their opponent $-i$ is playing a stationary strategy given by Π^{-i}. Then the player i would be able to calculate the value of an action $a \in \mathcal{A}_i$, given the state of the system is $s \in \mathcal{S}$, by calculating $Q^i(s,a)$ found in (2.3.3).

$$Q^i(s,a) = r^i(s,a) + \gamma \sum_{\tilde{s}\in\mathcal{S}} p(\tilde{s} \mid s,a) \max_{\tilde{a}^i \in \mathcal{A}^i} \mathbb{E}_{\tilde{a}^{-i}\sim\Pi^{-i}} \left\{ Q^i(\tilde{s},\tilde{a}) \right\}. \tag{2.3.3}$$

The function Q^i is known as player i's Q-function.

Fictitious Play

We present algorithms for fictitious play as found in Sayin et al. (2022). In fictitious play, agents try to learn both their opponent's strategy and the agent's own Q-function. Agents achieve this by smoothly updating their belief of the opponent's strategy based on a history of observed past plays. The agents then act greedily, in that they try only to maximize their own reward, based on their notion of the opponent's strategies.

We consider two different fictitious play algorithms for stochastic games: a model free and a model based method. A model based algorithm assumes that the agents know the reward functions and state transition functions within their environment. However, in a model free algorithm, the agents must learn the reward function and state transitions based off only their observations within the environment. Intuitively, a model based algorithm achieves quicker convergence to Nash equilibria since there is less to learn about the environment. However, in reality the parameters of the game might not be available to the agents and therefore this motivates the creation of model free algorithms.

Model Based Fictitious Play

Algorithm 1 shows the model based fictitious play algorithm as found in Sayin et al. (2022). The algorithm contains two tuneable sequences of learning parameters, α_c and β_c, the former controlling how quickly the beliefs of strategies change and the latter determining the speed at which Q-functions are updated.

Algorithm 1: Model Based Fictitious Play (Sayin et al., 2022)

Initialise: $\hat{Q}_0^1(s,a^1,a^2) = \hat{Q}_0^2(s,a^1,a^2) = 0$ for all $s \in \mathcal{S}, a^1 \in \mathcal{A}^1$, and $a^2 \in \mathcal{A}^2$, $\hat{\pi}_0^1(s)$ and $\hat{\pi}_0^2(s)$ uniform for all $s \in \mathcal{S}$, $\#s = 0$ for all $s \in \mathcal{S}$, k = 1

1 Observe the current state s_k

2 Take actions greedily,

$$a_k^1 \in \arg\max_{a^1 \in \mathcal{A}^1} \mathbb{E}_{a^2 \sim \hat{\pi}_{k-1}^2(s)} \left\{ Q_{k-1}^1(s_k, a^1, a^2) \right\}$$
$$a_k^2 \in \arg\max_{a^2 \in \mathcal{A}^2} \mathbb{E}_{a^1 \sim \hat{\pi}_{k-1}^1(s)} \left\{ Q_{k-1}^2(s_k, a^1, a^2) \right\}$$

3 Actions are observed by everybody.

4 Update belief of strategies for the state s_k by,

$$\hat{\pi}_k^1(s_k) = \hat{\pi}_{k-1}^1(s_k) + \alpha_{\#s}[a_k^1 - \hat{\pi}_{k-1}^1(s_k)]$$
$$\hat{\pi}_k^2(s_k) = \hat{\pi}_{k-1}^2(s_k) + \alpha_{\#s}[a_k^2 - \hat{\pi}_{k-1}^2(s_k)].$$

5 Update Q-functions as follows,

$$Q_k^1(s_k,a^1,a^2) = Q_{k-1}^1(s_k,a^1,a^2) + \beta_{\#s_k} \left(r^1(s_k,a^1,a^2) + \gamma \sum_{\tilde{s}} \mathbb{P}(\tilde{s} \mid s_k,a^1,a^2) \hat{v}_{k-1}^1(\tilde{s}) - Q_{k-1}^1(s_k,a^1,a^2) \right)$$

6
$$Q_k^2(s_k,a^1,a^2) = Q_{k-1}^2(s_k,a^1,a^2) + \beta_{\#s_k} \left(r^2(s_k,a^1,a^2) + \gamma \sum_{\tilde{s}} \mathbb{P}(\tilde{s} \mid s_k,a^1,a^2) \hat{v}_{k-1}^2(\tilde{s}) - Q_{k-1}^2(s_k,a^1,a^2) \right)$$

for all a^1, a^2 where,

$$\hat{v}_k^1(s) = \max_{a^1 \in \mathcal{A}^1} \mathbb{E}_{a^2 \sim \hat{\pi}_k^2(s)} \left\{ Q_{k-1}^1(s,a^1,a^2) \right\}$$
$$\hat{v}_k^2(s) = \max_{a^2 \in \mathcal{A}^2} \mathbb{E}_{a^1 \sim \hat{\pi}_k^1(s)} \left\{ Q_{k-1}^2(s,a^1,a^2) \right\}$$

Q-functions for other states $s \neq s_k$ remain constant.

7 $k := k + 1$.

Under the assumptions:

- $\alpha_c, \beta_c \in (0, 1)$ are non-increasing sequences.
- The sequences both have infinite sums, $\sum_c \alpha_c = \sum_c \beta_c = \infty$, and tend to zero $\lim_{c\to\infty} \alpha_c = \lim_{c\to\infty} \beta_c = 0$.
- The ratio of the sequences tends to zero $\lim_{c\to\infty} \beta_c/\alpha_c = 0$.

Sayin et al. (2022) prove that the strategies for each agent converge to strategies which form a Nash equilibrium $(\hat{\pi}^1_k, \hat{\pi}^2_k) \to (\hat{\pi}^1_*, \hat{\pi}^2_*)$, and the estimated Q-functions converge to the true Q-function $(\hat{Q}^1_k, \hat{Q}^2_k) \to (\hat{Q}^1_*, \hat{Q}^2_*)$ with probability one.

Model Free Fictitious Play

Algorithm 2 shows the model free fictitious play algorithm as found in Sayin et al. (2022). The algorithm contains two tuneable sequences of learning parameters, α_c and β_c, as in the model based fictitious play algorithm. There is also a parameter ϵ which controls the rate of exploration in the learning dynamics. With probability $1 - \epsilon$ the agents are greedy as in the model based algorithm, but with probability ϵ they now play an action uniformly at random to learn about its reward and effect on state transitions.

Algorithm 2: Model Free Fictitious Play (Sayin et al., 2022)

Initialise: $\hat{Q}^1_0(s,a^1,a^2) = \hat{Q}^2_0(s,a^1,a^2) = 0$ for all $s \in \mathcal{S}, a^1 \in \mathcal{A}^1$, and $a^2 \in \mathcal{A}^2$, $\hat{\pi}^1_0(s)$ and $\hat{\pi}^2_0(s)$ uniform for all $s \in \mathcal{S}$, $\#s = 0$ for all $s \in \mathcal{S}$, k = 1

1 Observe the current state s_k

2 Update Q-functions as follows,

$$\begin{aligned} Q^1_{k-1}(s_{k-1},a^1_{k-1},a^2_{k-2}) &= Q^1_{k-1}(s_{k-1},a^1_{k-1},a^2_{k-2}) \\ &\quad + \beta_{\#s_k}\left(r^1(s_{k-1},a^1_{k-1},a^2_{k-2}) + \gamma \hat{v}^1_{k-1}(s_k) - Q^1_{k-1}(s_{k-1},a^1_{k-1},a^2_{k-2})\right) \\ Q^2_{k}(s_{k-1},a^1_{k-1},a^2_{k-2}) &= Q^2_{k-1}(s_{k-1},a^1_{k-1},a^2_{k-2}) \\ &\quad + \beta_{\#s_k}\left(r^2(s_{k-1},a^1_{k-1},a^2_{k-2}) + \gamma \hat{v}^2_{k-1}(s_k) - Q^2_{k-1}(s_{k-1},a^1_{k-1},a^2_{k-2})\right) \end{aligned}$$

3

for all a^1, a^2 where,

$$\hat{v}^1_k(s) = \max_{a^1 \in \mathcal{A}^1} \mathbb{E}_{a^2 \sim \hat{\pi}^2_k(s)}\left\{Q^1_{k-1}(s,a^1,a^2)\right\} \text{ and } \hat{v}^2_k(s) = \max_{a^2 \in \mathcal{A}^2} \mathbb{E}_{a^1 \sim \hat{\pi}^1_k(s)}\left\{Q^2_{k-1}(s,a^1,a^2)\right\}$$

Agents act greedily with probability $1-\epsilon$ otherwise an action is drawn uniformly at random from their action space,

$$a^1_k = \begin{cases} a^1_* \in \arg\max_{a^1 \in \mathcal{A}^1} \mathbb{E}_{a^2 \sim \hat{\pi}^2_{k-1}(s)}\left\{Q^1_{k-1}(s_k,a^1,a^2)\right\} & \text{w.p. } (1-\epsilon) \\ u^1 \sim \mathcal{U}(\mathcal{A}^1) & \text{w.p. } \epsilon. \end{cases}$$

$$a^2_k = \begin{cases} a^2_* \in \arg\max_{a^1 \in \mathcal{A}^2} \mathbb{E}_{a^1 \sim \hat{\pi}^1_{k-1}(s)}\left\{Q^2_{k-1}(s_k,a^1,a^2)\right\} & \text{w.p. } (1-\epsilon) \\ u^2 \sim \mathcal{U}(\mathcal{A}^2) & \text{w.p. } \epsilon. \end{cases}$$

4 Actions are observed by everybody.

5 Update belief of strategies for the state s_k by,

$$\hat{\pi}^1_k(s_k) = \hat{\pi}^1_{k-1}(s_k) + \alpha_{\#s}[a^1_k - \hat{\pi}^1_{k-1}(s_k)] \text{ and } \hat{\pi}^2_k(s_k) = \hat{\pi}^2_{k-1}(s_k) + \alpha_{\#s}[a^2_k - \hat{\pi}^2_{k-1}(s_k)].$$

Strategies for other states $s \neq s_k$ remain constant.

6 $k := k+1$.

Under the assumptions:

- $\alpha_c \in (0,1)$ is a non-increasing sequence and $\beta \in (0,1)$ is a monotonically decreasing sequence.
- The sequences both have infinite sums, $\sum_c \alpha_c = \sum_c \beta_c = \infty$, but finite sum of squares $\sum_c \alpha_c^2 = \sum_c \beta_c^2 < \infty$, and tend to zero $\lim_{c\to\infty} \alpha_c = \lim_{c\to\infty} \beta_c = 0$.
- The ratio of the sequences satisfies the condition that for any $m \in (0,1]$ we have $\lim_{c\to\infty} \beta_{\lfloor mc \rfloor}/\alpha_c = 0$.

Sayin et al. (2022) prove that beliefs on strategies and Q-function converge to a near equilibrium and the equilibrium Q-functions with an approximation linear in the exploration probability $\epsilon > 0$, almost surely. Therefore, with probability one we have

$$\limsup_{k\to\infty} |U^i(\pi^i, \pi_k^{-i}) - U^i(\hat{\pi}_k^i, \hat{\pi}_k^{-i})) \leq 2\epsilon D \frac{(1+\gamma)^2}{\gamma(1-\gamma)^3}$$

and

$$\limsup_{k\to\infty} |\hat{Q}_k^i(s,a) - Q_i^*(s,a)| \leq \epsilon D \frac{1+\gamma}{(1-\gamma)^2}$$

where

$$D = \frac{1}{1-\gamma} \sum_i \max_{(s,a)} |r^i(s,a)|$$

Q-Learning

We present the algorithm for Q-Learning in stochastic games as published by Sayin et al. (2021). In the fictitious play learning dynamics, it is assumed that the agents both observe the action of the opponent and keep a history of those actions. However, in certain applications this assumption may be unrealistic. For example, with the problem of border patrol, the patroller would need to have a perfect observation of the smugglers' actions at every location. However, the patroller might only be able to

observe the economic impact and therefore infer the effect of the smugglers' action on their reward.

In the Q-Learning algorithm, the agents do not have the knowledge that there is an opponent in the game. Therefore, rather than looking to find the Q-functions of the game, we instead look to define a function which depends only upon the action of that player. In order to do this, we denote the Q-function as used previously by the global Q-function of the game, whilst defining a local q-function for each player to only depend on the state and their action. Thus, we can find of a local q-function for a player i given a state s and action a_i as,

$$q_i(s, a_i) = \mathbb{E}_{a_{-i} \sim \pi_{-i}}[Q^i(s, a_i, a_{-i})]$$

where Q^i is player i's Q-function as defined in (2.3.3). If every player picked their best response by maximizing their local q-function, then the learning dynamics could get stuck playing suboptimal actions. Therefore, there is motivation to consider smoothed best responses, where there is some probability assigned to each action according to how it compares to the best response. Using smoothed best responses allows the learning dynamics to converge to the Nash equilibrium, rather than potentially getting caught in cyclic behavior.

We define the smoothed best response for the player i given their estimated local q-function $\hat{q}$, the current state s_k, and $\beta_{\#s_k}$ the number of times the game has been in state s_k to be $\overline{\text{Br}}(\hat{q}, \tau_{\#s_k}) \in \Delta(\mathcal{A})$ where $\tau_{\#s_k}$ is a temperature parameter. The authors of Sayin et al. (2021) use the smoothed best response defined as,

$$\overline{\text{Br}}(\hat{q}, \tau_{\#s_k}) = \underset{\mu \in \Delta(\mathcal{A})}{\arg\max} \{\mu \cdot \hat{q} + \tau_{\#s_k} \nu(\mu)\}$$

where ν is a smooth and strictly concave function with unbounded gradient at the boundary of the simplex $\Delta(\mathcal{A})$ which was first introduced by Fudenberg and Levine

(1998). The temperature parameter $\tau > 0$ determines how smoothed the best response will be. The choice of choosing a smoothed best response in this fashion guarantees that there will exist a unique smoothed best response π. Under the choice of ν to be,

$$\nu(\mu) = -\sum_{a \in \mathcal{A}} \mu(a) \log(\mu(a))$$

results in the simplification of,

$$\pi(a) = \frac{\exp(\hat{q}(a)/\tau_{\#s_k})}{\sum_{\tilde{a}} \exp(\hat{q}(\tilde{a})/\tau_{\#s_k})} > 0$$

Given the above choice for the smoothed best response, the algorithm for decentralized Q-learning from Sayin et al. (2021) is found in Algorithm 3.

Under the assumptions:

- Given any pair of states (s, s_0) and any infinite sequence of actions, s_0 is reachable from s with some positive probability within a finite number, n, of stages.
- The sequence $\{\tau_c\}_{c>0}$ is non-increasing and satisfies $\lim_{c\to\infty}(\tau_{c+1} - \tau_c)/\alpha_c = 0$ and $\lim_{c\to\infty} \tau_c = 0$.
- The step size $\{\alpha_c\}_{c>0}$ satisfies $\sum_{c=1}^{\infty} \alpha_c^{2-\rho} < \infty$ for some $\rho \in (0, 1)$.
- There exists $C, C' \in (0, \infty)$ such that $\alpha_c^{\rho} \exp(4D/\tau_c) \leq C'$ for all $c > C$.

Sayin et al. (2021) prove that the estimated value functions tend to the true value function under Nash equilibrium,

$$\lim_{k\to\infty} |\hat{v}_{s,k}^i - v_{\pi^*}^i(s)| = 0.$$

Algorithm 3: Decentralized Q-Learning

Initialise: $\hat{q}^1_{s,0}, \hat{q}^2_{s,0} = (0, \ldots, 0)$, $\hat{v}^1_{s,0}, \hat{v}^2_{s,0} = 0$, $\#s = 0$ for all $s \in \mathcal{S}$

1 Observe the current state s_k

2 Update the local Q-function estimate for previous state s_{k-1} and previous local action b_{k-1} and $\boldsymbol{a}_{k-1}$ respectively.

$$\hat{q}^1_{s_{k-1},k}[a^1_{k-1}] = \hat{q}^1_{s_{k-1},k-1}[a^1_{k-1}] + \overline{\alpha}^1_{k-1}\left(r^1_{k-1} + \gamma \hat{v}^1_{s_k,k-1} - \hat{q}^1_{s_{k-1},k-1}[a^1_{k-1}]\right)$$
$$\hat{q}^2_{s_{k-1},k}[a^2_{k-1}] = \hat{q}^2_{s_{k-1},k-1}[a^2_{k-1}] + \overline{\alpha}^2_{k-1}\left(r^2_{k-1} + \gamma \hat{v}^2_{s_k,k-1} - \hat{q}^2_{s_{k-1},k-1}[a^2_{k-1}]\right)$$

where,

$$\overline{\alpha}^1_{k-1} = \min\left\{1, \frac{\alpha_{\#s_{k-1}}}{\overline{\pi}^1_{k-1}[a^1_{k-1}]}\right\}$$
$$\overline{\alpha}^2_{k-1} = \min\left\{1, \frac{\alpha_{\#s_{k-1}}}{\overline{\pi}^2_{k-1}[a^2_{k-1}]}\right\}$$

3 Increment state counter: $\#s_k := \#s_k + 1$.

4 Take actions $a^1_k \sim \overline{\pi}^1_k$ and $a^2_k \sim \overline{\pi}^2_k$ where,

$$\overline{\pi}^1_k[a^1] = \frac{\exp\left(\hat{q}^1_{s_k,k}[a^1]/\tau_{\#s_k}\right)}{\sum_{\tilde{a}^1 \in \mathcal{A}^1} \exp\left(\hat{q}^1_{s_k,k}[\tilde{a}^1]/\tau_{\#s_k}\right)}$$
$$\overline{\pi}^2_k[a^2] = \frac{\exp\left(\hat{q}^2_{s_k,k}[a^2]/\tau_{\#s_k}\right)}{\sum_{\tilde{a}^2 \in \mathcal{A}^2} \exp\left(\hat{q}^2_{s_k,k}[\tilde{a}^2]/\tau_{\#s_k}\right)}$$

5 Collect the local rewards r^1_k and r^2_k and update value function estimates according to,

$$\hat{v}^1_{s_k,k+1} = \hat{v}^1_{s_k,k} + \beta_{\#s_k}\left[\bar{\pi}^1_k \cdot \hat{q}^1_{s_k,k} - \hat{v}^1_{s_k,k}\right]$$
$$\hat{v}^2_{s_k,k+1} = \hat{v}^2_{s_k,k} + \beta_{\#s_k}\left[\bar{\pi}^2_k \cdot \hat{q}^2_{s_k,k} - \hat{v}^2_{s_k,k}\right]$$

Chapter 3

Border Patrol

There is a significant operations research literature on patrol problems that focuses on modelling real-world situations. In this chapter, we present three different border patrol problems in the literature in detail, before giving a brief but wider overview of the area as a whole.

3.1 Patrolling the Los Angeles International Airport

Pita et al. (2008) introduced a model for patrolling the Los Angeles International Airport (LAX) which was successfully deployed. Given the size of the airport, with the number of checkpoints and terminals which could potentially be monitored, it is a large issue of how to generate schedules for security teams. The authors developed a model for protecting the airport using a Bayesian Stackelberg game, and provided the fastest known way to solve the game. Their model allowed the user to adjust the schedule if extra constraints occurred, and also warned the user if these changes had sufficiently degraded the schedule generated.

In a Stackelberg game, there are two players: a leader and a follower. The leader

first commits to a strategy, which could be mixed, and then the follower observes the strategy and chooses a strategy to optimize their reward. Bayesian Stackelberg games extend the framework by allowing for a set of N agents, where agent n is one of a given set of types θ_n. The Bayesian Stackelberg games considered by Pita et al. (2008) consider two agents, the leader and follower, where the leader only has one type, but there are multiple possible follower types. Each follower type corresponds to a different adversary who might try to breach the security of the airport, with different objectives and payoffs for their actions.

The model introduced by Pita et al. (2008) has the following framework, which can be solved as a mixed-integer quadratic program. The leader's policy is denoted by x, which is a probability distribution over the pure strategies $i \in X$ with x_i being the probability an action i is chosen. There is a set L of possible follower types, with p_l being the probability the follower is of type $l \in L$. A follower of type l has their policy denoted as q^l, which is a probability distribution over the pure strategies $j \in Q$ with q^l_j being the probability an action j is chosen. The payoff matrices to the leader and follower, dependent on the follower being type l, are given by R^l and C^l respectively. The leader's strategy can then be found by solving the following, where M is some large positive integer.

$$
\begin{aligned}
\max_{x,q,a} & \sum_{i \in X} \sum_{l \in L} \sum_{j \in Q} p^l R^l_{ij} x_i q^l_j \\
\text{s.t.} & \sum_{i \in X} x_i = 1 \\
& \sum_{j \in Q} q^l_j = 1 \\
& 0 \leq \left(a^l - \sum_{i \in X} C^l_{ij} x_i \right) \leq \left(1 - q^l_j\right) M \\
& x_i \in [0,1] \\
& q^l_j \in \{0,1\} \\
& a \in \mathbb{R}
\end{aligned}
$$

The authors linearize the quadratic programming problem by using a change of variables, and show that the resulting mixed integer linear program can then be solved with existing efficient integer programming computing packages.

There are a number of other models in the literature which use Stackelberg games to model problems of border patrol.

A similar model is considered by Shieh et al. (2012), where the authors present a game theoretic model of the protection of ports in the United States. In particular, they develop a system which has been deployed by the United States Coast Guard for the protection of the port of Boston. The authors use a Stackelberg game to model the interaction between an attacker with the defender. A key feature of the model is they do not assume perfect rationality of the attacker, since this may not be realistic, and instead use a quantal response model of their behavior.

Yang et al. (2014) also develop a Stackelberg game model to tackle the issue of illegal poaching. The authors develop the model in a joint effort with the Queen Elizabeth National Park in Uganda to help improve the wildlife ranger patrols. Using data

collected from previous poaching, a behavioral model is trained to learn the poachers' decision-making process.

Brown et al. (2006) look at how critical infrastructure can be made more resilient against attacks from terrorists. The authors use a Stackelberg game to model the problem, and consider a number of real life examples including the US Strategic Petroleum Reserve, the US Border Patrol at Yuma, Arizona, and an electrical transmission system.

3.2 Patrolling games model

Alpern et al. (2011) introduced a different framework to model the patrolling of a border. The game formulated by the authors is a two player zero-sum game between an attacker and a patroller. The attacker wins if they make a successful attack, and loses otherwise. The patroller's outcomes are the opposite, since the game is zero-sum. There is a finite time horizon of length T denoted by $\mathcal{T} = \{0, 1, \ldots, T\}$, and the game is played on a graph Q which consists of n nodes $\mathcal{N}$ joined by edges $\mathcal{E}$. The attacker can choose a node to attack, and then their attack lasts for some m time steps. During the time when the attack is occurring, it is stopped by the patroller if they visit that node of the graph. Attacks are always caught, and stopped instantaneously. The attacker's pure strategies consist of the pairs $[i, I]$ where $i \in \mathcal{N}$ is the attack node and $I = \{\tau, \tau + 1, \ldots, \tau + m - 1\}$ is the attack interval. A pure strategy for the patroller is a walk $w : \mathcal{T} \to Q$, which is called a patrol. There are two cases of the game: either a one-off game or a periodic game. The one-off game is as described, and the attack must be completed by time T for the attacker to win. In the periodic game, we restrict the patroller's pure actions to be cycles of length T in order to have the patrols join up. Attacks are now not required to be completed by time T, and if an attack starts at time τ, it finishes at the time point $\tau + m \mod T$.

The main contributions of Alpern et al. (2011) are discussing patrolling on special

classes of graphs. There are three classes of note: Hamiltonian, bipartite and line graphs.

A Hamiltonian graph is defined as a graph where there exists a cycle which visits every node exactly once. Alpern et al. (2011) prove that for any Hamiltonian graph with n nodes, if the attack lasts for m time steps then the value of the one-off game is m/n and the value of the periodic game is bounded above by m/n. Furthermore, in the periodic game the value is exactly m/n if $T = kn$ for some $k \in \mathbb{Z}$, or if $T = kn + \sigma$ for some $0 < \sigma < n$, as $k \to \infty$ the value of the game tends to m/n.

A bipartite graph is a graph in which there exists no cycle of odd length. Therefore, in bipartite graphs the nodes can be partitioned into two sets $\mathcal{A}$ and $\mathcal{B}$ such that every edge is from a node in $\mathcal{A} = \{\alpha_1, \ldots, \alpha_a\}$ to a node in $\mathcal{B} = \{\beta_1, \ldots, \beta_b\}$, where $a \leq b$. If every node in $\mathcal{A}$ is connected by an edge to every node in $\mathcal{B}$, then the graph is called the complete bipartite graph $K_{a,b}$. The authors prove that for any bipartite graph, with nodes partitioned as mentioned, in both the one-off game and the period game the value is bounded above by $m/(2b)$. Moreover, the bound is tight if the graph is the complete bipartite graph $K_{a,b}$ in the one-off game, or if it is the complete bipartite graph $K_{a,b}$ and $T = 2kb$ in the periodic game for some $k \in \mathbb{Z}$. Additionally, if $T = kn + \sigma$ for some $k, \sigma \in \mathbb{Z}$ such that $0 < \sigma < n$ then as $k \to \infty$ the value of the periodic game tends to $m/(2b)$.

The line graph L_n is a graph of n nodes such that there are only edges between nodes k and $k+1$ for $1 \leq k \leq n-1$. Alpern et al. (2011) prove that for the line graph L_n, the value of the one-off game is $m/2(n-1)$ and the value of the periodic game is bounded above by $m/2(n-1)$. Furthermore, in the periodic game the value is exactly $m/2(n-1)$ if $T = kn$ for some $k \in \mathbb{Z}$, or if $T = kn + \sigma$ for some $0 < \sigma < n$ we get as $k \to \infty$ the value of the game tends to $m/2(n-1)$. Papadaki et al. (2016) and Alpern et al. (2019) analyze the line graph case in further detail.

Alpern et al. (2022a) extend the model of Alpern et al. (2011) to both continuous

space and time. The patroller now moves around a network at a unit speed, and wins if they intercept the attack being carried out by the attacker. The arcs of the network are not assumed to have equal length, and we define μ to be the sum of all arc lengths. The girth g of the network Q is defined as the minimum length of a circuit in Q (or $g = \infty$ if no circuit exists). The authors prove that for any network without leaf arcs, if $\alpha \leq g$ the value of the game is equal to α/μ. Furthermore, the authors can give properties about the strategies played by either side. For the attacker, any uniform attack strategy is optimal. For the patroller, a tour is defined which covers every arc twice and no arc is traversed consecutively in opposite direction. The analysis is then extended to general graphs, which can include leaf nodes. The generalized girth g^* of a network Q is defined to be the smaller between either the shortest circuit length of Q and twice the length of the shortest leaf arc. The authors prove that for a network with $l \geq 0$ leaf nodes and generalized girth g^*, if $\alpha \leq g^*$ the value of the game is equal to,

$$\frac{\alpha}{\mu + l\alpha/2}$$

Optimal strategies for the attacker and patroller are also once more defined by the authors.

There are a number of other extensions to Alpern et al. (2011) considered in the literature. Lin et al. (2013) consider how the problem changes if instead of attacks taking a deterministic length of time to complete, the player's only know a distribution which it is drawn from. A further extension is made by Lin et al. (2014) by adding the chance for the patroller to overlook an attack taking place. McGrath and Lin (2017) look at what happens when the time taken to traverse an edge in the graph is non-zero. In Alpern et al. (2022b), the authors consider an extension where the attacker can observe the presence of the patroller, and based off the information choose to delay the time of their attack.

3.3 Travelling inspector model

The travelling inspector model was first introduced by Filar (1985), where a game is played in which an inspector checks if regulations are being followed at different plants. The plants are in different geographical locations, meaning the travel from one to another is not trivial, and each controlled by some inspectee. The inspectee benefits from not following the regulations, for example by increasing their plant's performance, but faces a fine if they get caught doing so. The inspector's aim is to minimize a combination of the costs inflicted by both the undetected violations and their travel. Filar introduces the travelling inspector model as a single-controller stochastic game with the following structure:

- There are S plants in different locations.
- There is one inspector who can perform only one inspection at a time.
- The inspector travels to a location, or remains still, and performs an inspection there.
- The inspectees only know the last inspected site.
- The inspector minimizes a combination of the costs inflicted by both the undetected violations and their travel.
- The number of inspection periods can either be finite or infinite, but is known to all players.

Each inspectee has levels with which they can violate the regulations. Filar denotes the set of possible actions for inspectee p by $V(p)$ for $p \in [n]$. Each inspectee receives a reward, depending on only their action $v_p \in V(p)$, the inspector's action i, and the state s, denoted by $r_p(v_p, i, s)$. The inspector's reward $r_I(v_1, \ldots, v_p, i, s)$ depends on their action, the actions of all smugglers and the state.

Filar and Schultz (1986) prove results about smuggler aggregation and its consequences. The travelling inspector game, as introduced above, is given as an $S + 1$ player game. However, if we assumed the S inspectees to be a single player, for example if they were controlled by some central decision maker, it is of interest how the equilibria would change. Filar showed that the sets of Nash equilibria between the two games in fact coincide. The result is both interesting from a policy perspective, but furthermore allows for easier computation of the equilibria from the wider range of literature on two player games.

3.4 Search Games

A related problem to border patrolling is the area of search problems, in which a hidden item needs to be found by a searcher. Applications include finding bombs, looking for missing planes or ships, and an intruder of a protected area.

The classic search problem is one where there is a single decision maker, the searcher, that tries to find the hidden item. Whilst the item may not necessarily be stationary, it does not have any ability to control its movement. The problem is called a one-sided search problem, and applications include trying to find a wrecked plane or ship. An overview of one-sided search problems can be found in Stone (1976).

However, there are many applications where either the item may be moving intelligently, for example an intruder, or may have been planted adversarially, as with a hidden bomb. There are now two decision makers, the searcher, and the hider, therefore these are called two-sided search problems. An overview of two-sided search problems can be found in Alpern and Gal (2006).

3.5 Wider Literature

In addition to the selected areas already discussed, there is a breadth of wider research in the area of border patrol style problems. Here we cover some of the different problems which have been considered in the literature.

The papers of Baston and Bostock (1991) and Garnaev (1994) discuss a stochastic game model for an inspection problem, applications of which include patrolling problems. Both papers consider a single smuggler, with a constraint on the number of times the patroller can attempt to capture the smuggler. The models also include a limited amount of time in which the smuggler can be caught.

Bier et al. (2007) consider a model where the defender must allocate resources to a number of locations, whilst the attacker chooses a location to attack. The paper considers which strategies for the two sides result in equilibrium.

Grant et al. (2020) examine a patrolling problem along a border where there are many small attempts to smuggle items. The adversaries are assumed to act randomly, rather than intelligently, with applications including photographing wildlife.

Lindelauf et al. (2009) consider the optimal communication structure of terrorist organizations, looking at the tradeoff between secrecy and operational efficiency. The authors model the problem as a game theoretic bargaining problem and find equilibria under a number of different assumptions.

Richard (1972) studies the daily patrol patterns of a police officer in the United States. A number of topics are considered including the models of police response time, preventative patrol effectiveness, workload distribution, dispatch delays, intersector cooperation, and a number of other performance measures.

Ruan et al. (2005) consider patrolling units that respond to calls for service. The locations have different levels of priority, and varying rates at which incidents occur. The authors develop a Markov decision process framework with a novel learning algorithm to find patrol routes.

The papers of Ruckle (2001) and Kikuta and Ruckle (2002) look at an accumulation game, where a hider distributes some material over a number of locations. A searcher then can choose a subset of the locations to visit, confiscate any material found, and confiscate it. There is a win-loss outcome to the game, depending on if the amount of material remaining exceeds some threshold known before the game to both sides.

Washburn and Wood (1995) consider a two-sided problem of network interdiction. A single evader attempts to traverse between two nodes in a network, whilst the patroller sets up an inspection point along one of the arcs in the network. The evader attempts to pick the shortest possible path without being caught, while the patroller tries to pick the arc to maximize their probability of capture. The authors construct a linear program to solve the game, and analyze the complexity of the problem. There are many extensions to the problem, a number of which are detailed in the overview by Smith and Song (2020).

Sack and Urrutia (1999) look at computational geometry, an application of which is the protection of galleries containing expensive paintings.

Chapter 4

A model of cooperation and communication

Chapter 4 considers a one-shot game between a patroller and a group of smugglers. The contribution of the chapter to the literature is its novel consideration of how communication and cooperation can be modelled, and furthermore, how they affect strategies at equilibria. This chapter is currently under review for publication in Operations Research.

This chapter is currently under review at Operations Research.

4.1 Introduction

The question of how to patrol a border effectively is a fundamental problem facing government organizations worldwide. A specific issue regarding borders is the illegal trafficking of items across them, with smugglers attempting to evade being captured as they pass through. Notable examples include drug trafficking through Europe (Baniya, 2023), oil smuggling (Savage and Bergman, 2023), and the illicit trade of wildlife (Freedman, 2022). Due to constraints on resources, it is usually infeasible to protect an entire border at once. Therefore, it is necessary to study how borders can be best defended

with limited resources.

Previous research has shown how the defensive strategies adopted by patrollers can directly affect the behavior of smugglers. Chalmers et al. (2009) discuss how one of the main aims of law enforcement targeting the trafficking of illicit drugs is the need for smugglers to increase the market price in order to compensate for increasing the risks they face. Rhodes et al. (2000) studied the impact of law enforcement policy on the market price of illicit drugs such as cocaine, heroin, marijuana and methamphetamine and concluded that without the interdiction of law enforcement, the market prices for these drugs would likely be many times lower.

In this chapter we make the assumption that as the overall quantity of items successfully smuggled increases, the smugglers receive diminishing marginal returns. This assumption can be justified using examples from the economics literature. Becker (1968) considers a general model of crime and punishment and assumes that offenders eventually receive diminishing marginal returns. In the context of smuggling, Sheikh (1974) and Norton (1988) also include notions of diminishing marginal returns. Both papers seek to build models for the smuggling of legal items through borders by companies in order to avoid taxation of their goods.

The inclusion of diminishing marginal returns raises non-trivial questions about how smugglers should collaborate with each other. Some previous studies have investigated the effects of communication and cooperation between smugglers. Politi (1997) states the importance of drug trafficking for organized crime, with two of the main incentives being the economic resources generated and the transnational networks created and sustained by the activity. Bichler et al. (2017) look into the structure of drug supply networks created by organized crime, analyzing them using techniques from social network analysis. In this chapter we use a game theoretical model to study the incentives for smugglers to cooperate and/or communicate with each other and the benefits obtained through collaboration, thereby demonstrating the importance of organizations

for drug trafficking.

A number of previous works have analyzed patrolling problems from a game theoretic perspective. Alpern et al. (2011) introduced a patrol game where a patroller attempts to thwart a single attack from a smuggler, with the attack taking a known deterministic time to complete. The game is zero-sum, with the smuggler winning if the attack is successful and losing if they are stopped by the patroller. Subsequent models analyzed include those of Lin et al. (2013), which introduces a non-deterministic time for attacks to be completed, Lin et al. (2014), which considers non-perfect detection of the smuggler by the patroller, and Papadaki et al. (2016); Alpern et al. (2019) which focus on applications to the more specific setting of border patrol.

The papers mentioned above all feature assumptions that make them incompatible with the model studied in this chapter. The game theoretical model in our chapter does not assume that a successful attack by a smuggler simply results in a loss for the patroller. Instead, we incorporate a more detailed payoff structure which takes into account the quantity of items trafficked, which is motivated from both the patroller's and the smugglers' perspective. The patroller must accept that they cannot always prevent all smuggling, and consequently, their payoff should be decreasing in the number of illicit items smuggled. Furthermore, the smugglers should not simply win the game if some illicit items are successfully smuggled. We present a payoff function for the smugglers which takes into account both the individual rewards for smuggling items but also depends on the total quantity of items smuggled, to take into account the diminishing returns received.

Models related to ours are considered in Filar and Schultz (1986) and Darlington et al. (2023). Both papers consider stochastic game models in which a patroller attempts to prevent a group of smugglers from attacking the border, but these papers also consider sequential problems with multiple time steps and the patroller must pay a cost to change the location that they are defending. Whilst we are only considering

a single time-step (or 'one-shot') problem, we introduce extra complexity through the structures of the payoff functions. The smuggler payoffs in Filar and Schultz (1986) and Darlington et al. (2023) have the property that one smuggler's payoff is independent of every other smuggler's action. This property implies that the equilibria in games with selfish and cooperative smugglers are equivalent, which leads to significant simplifications when finding equilibrium strategies. When diminishing marginal returns are included, it is no longer the case that one smuggler's payoff is independent of other smugglers' actions, and we must therefore establish new methods to find equilibria.

Our model aims to yield useful insights that can assist a patroller to choose the best strategy when defending a border against smugglers. Moreover, these insights have potential design implications, as the patroller could potentially exert influence over the values of the model parameters in realistic settings. For example, we assume that the cost to a smuggler of being caught consists of both the amount of revenue lost and also the penalty imposed upon them by the patroller. This penalty could be chosen (possibly within constraints) by the patroller in order to either incentivize or deter smugglers from attempting to attack. Kleiman and Kilmer (2009) present a study of the effects of choosing different levels of deterrent in a simple model that does not include dependence between the rewards to smugglers and the amount of crime occurring.

The main contributions of our chapter are as follows:

- We introduce a new model for patrolling a border which takes into consideration diminishing marginal returns for illicit items being trafficked by a group of smugglers.
- We consider three different cases of smuggler behavior: selfish smugglers without communication, selfish smugglers with communication and cooperative smugglers.

- We prove analytically that in all cases of smuggler behavior, any equilibrium strategy requires the smugglers to attack each location with the same probability. Furthermore, in two of the three cases, we show that there can only exist an equilibrium when the patroller defends each location with the same probability.

- We detail all possible equilibria that exist in two of the three cases of smuggler behavior. Moreover, after placing an additional assumption on the patroller's behavior, we describe all possible equilibria in the remaining case.

- We prove the surprising result that cooperation, and sometimes communication, between the smugglers results in fewer illicit activities. Moreover, both cooperation and communication increase the value of the game to the smugglers.

- We consider how the penalty applied to smugglers who get caught affects both their actions and their payoffs. Choosing the deterrent is an important societal consideration to balance the consequences for the criminal with the impact of crime on the population as a whole.

The rest of the chapter is organized as follows. In Section 4.2 we formulate the game theoretical model for border patrol. In Section 4.3 we present the different definitions of equilibria that we wish to consider for the players in the model. Section 4.4 discusses the smugglers' strategy and proves certain properties of equilibrium strategies that apply to all behavior cases. Section 4.5 considers each case independently, and provides analytical proofs regarding the structures of the various equilibria. In Section 4.6 we prove additional results based on comparisons between the equilibria in different cases. Section 4.7 includes examples to demonstrate the impact of the model parameters on the existence and nature of equilibrium strategies. Finally, Section 4.8 provides our concluding remarks.

4.2 Model Description

We consider a one-shot simultaneous game played between a patroller and n smugglers. Each smuggler is fixed at a unique and discrete location, meaning that the border is comprised in total of n locations labelled $1, 2, \ldots, n$. The patroller chooses to defend one of the locations, whilst each smuggler can choose whether or not to attack their respective location. We will use the notation $[n] = \{1, \ldots, n\}$ to denote the set of locations. As a result, the patroller's action space is denoted by $\mathcal{A}_{pat} = [n]$ and smuggler i's action space is given by $\mathcal{A}_{smug(i)} = \{0, 1\}$ for $i = 1, \ldots, n$. We denote the action chosen by the patroller by $b \in \mathcal{A}_{pat}$ and the action chosen by smuggler i by $a_i \in \mathcal{A}_{smug(i)}$. The vector $\boldsymbol{a} = (a_1, \ldots, a_n)$ represents the actions taken by all the smugglers.

We begin by describing how the smuggler rewards depend on the actions of all players. When a smuggler attacks, they incur a fixed cost of $C > 0$ if the patroller defends their location. If the smuggler attacks and their location is not defended by the patroller, then the attack is successful. When a smuggler makes a successful attack, they receive a positive reward, which depends on the total number of successful attacks. The total number of successful attacks is the total number of attacks at locations other than b, which for the smuggler actions $\boldsymbol{a}$ we denote by

$$\alpha_b(\boldsymbol{a}) = \sum_{j \neq b} a_j.$$

Given that $\alpha_b(\boldsymbol{a})$ successful attacks occur, the reward to each smuggler who makes a successful attack is given by $f(\alpha_b(\boldsymbol{a}))$, where the function f is specified within the model. We assume that $f : [n] \to \mathbb{R}$ is a positive and strictly decreasing function. Whilst in reality the reward could be negative (due to costs outweighing the benefits of smuggling), this would lead to uninteresting solutions in our model. The assumption that the reward for each successful attack is decreasing in the total number of attacks is motivated by studies which consider the price elasticity of supply for illicit items, such

as Rhodes et al. (2000).

Consequently, we define smuggler i's reward function as follows;

$$r_{smug(i)}(b, a_1, \ldots, a_i, \ldots, a_n) = \begin{cases} -a_i C, & b = i, \\ a_i f(\alpha_b(\boldsymbol{a})), & b \neq i. \end{cases} \tag{4.2.1}$$

We place one additional assumption on the cost parameter C and the reward function f. For any positive number x of successful attacks, we assume that the total reward obtained from these attacks, $xf(x)$, is strictly greater than the corresponding reward when one of the attacks is instead defended against, $(x-1)f(x-1) - C$. That is:

$$xf(x) > (x-1)f(x-1) - C \quad \forall \quad x > 0. \tag{4.2.2}$$

This assumption is necessary in order to avoid degenerate cases where a group of smugglers would be better off making sure that one of their attacks is stopped, rather than making all of their attacks successfully. Note that this assumption does not imply that

$$xf(x) > (x-1)f(x-1)$$

and therefore there can be cases where, for a group of smugglers, the total reward is increased by reducing the number of successful attacks that are made.

Since each smuggler's reward depends on the actions of the other smugglers, there is an intuitive incentive for them to cooperate rather than acting selfishly. Furthermore, even if the smugglers act selfishly, there could still be some centralizing agent outside of the game who coordinates their attacks to increase profits. These considerations provide motivation for studying the effects of cooperation and communication between smugglers on equilibrium solutions.

We now present the patroller's reward function. If a patroller defends a location that is being attacked by a smuggler, they receive a fixed reward of $c > 0$. We do not

assume that the game is zero-sum and therefore c and C have no relation. The patroller also incurs a cost that depends on the total number of successful attacks that they are unable to stop. We assume that there is a linear relationship between the number of attacks x that are not defended against, and the cost $g(x)$ to the patroller. It is further assumed that the linear function g is both positive and increasing. Governments are normally risk-neutral in their decision making (Stewart et al. (2011)), and therefore the linear assumption of their utility is justified. A detailed discussion of the societal cost of crime can be found in Wickramasekera et al. (2015). The patroller's reward function can be expressed as

$$r_{pat}(b, \boldsymbol{a}) = c \cdot a_b - g(\alpha_b(\boldsymbol{a})). \tag{4.2.3}$$

As g is a linear function, we can also write $g(x) = g_1 + g_2 x$ when needed. The constants g_1 and g_2 can be respectively interpreted as the constant cost of patrolling and the cost to the patroller for each individual attack that they fail to defend.

In this section we have detailed the actions available to the players in the game and how the payoffs depend on the choices of action. The next section describes the different cases for that we wish to consider for the smugglers' behavior.

4.3 Defining settings of cooperation and communication

In this section we formally define each of the cases of smuggler behavior and specify the conditions for equilibria in the game. The three different cases we consider are: (i) selfish smugglers without communication (abbreviated as S-NC), (ii) selfish smugglers with communication (abbreviated as S-C) and (iii) cooperative smugglers (abbreviated as CP). When smugglers are acting selfishly they act to maximize their own expected rewards. If they are not communicating, then they must choose actions independently

of each other. On the other hand, the ability to communicate allows them to observe the planned actions of other smugglers, and consequently their actions can depend on each other. When the smugglers are cooperative they aim to maximize the sum of their individual expected rewards, rather than maximizing individual rewards.

We define a strategy for a player to be a probability distribution over their action space. We use $\Delta(A)$ to denote the set of probability distributions with set A as their support. In all three of the cases introduced, the patroller chooses a strategy $\boldsymbol{q} \in \Delta(\mathcal{A}_{pat}) = \Delta([n])$. Given the strategy $\boldsymbol{q}$, the probability that a location k is defended is given by $q_k = \mathbb{P}(b = k)$.

4.3.1 Selfish - No Communication (S-NC)

In the case of selfish smugglers without communication, we aim to find a Nash equilibrium (as first introduced by Nash (1951)) between the patroller and the n smugglers. The patroller chooses a strategy over $\Delta([n])$ as previously described, whilst each smuggler i chooses a strategy $\boldsymbol{\pi}_i = (\pi_i^0, \pi_i^1) \in \Delta(\mathcal{A}_{smug(i)}) = \Delta(\{0, 1\})$. The probabilities of smuggler i taking the actions 0 and 1 are then represented by π_i^0 and π_i^1 respectively.

The expected rewards for all players, given a strategy for each player, can then be found by taking the expectations of their reward functions over the actions drawn at random from the respective strategies. The patroller's expected reward is

$$R_{pat}(\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n) = \mathbb{E}_{b\sim\boldsymbol{q}, a_1\sim\boldsymbol{\pi}_1, \ldots, a_n\sim\boldsymbol{\pi}_n}\left[r_{pat}(b, a_1, \ldots, a_n)\right]$$

and the expected reward for smuggler i is

$$R_{smug(i)}(\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n) = \mathbb{E}_{b\sim\boldsymbol{q}, a_1\sim\boldsymbol{\pi}_1, \ldots, a_n\sim\boldsymbol{\pi}_n}\left[r_{smug(i)}(b, a_1, \ldots, a_n)\right].$$

The strategies $\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n$ are a Nash equilibrium if and only if no player has an incentive to deviate to another strategy. A player has an incentive to deviate to another

strategy if they can receive a strictly greater expected reward by playing it, assuming the strategies of all other players remain the same. Therefore, the patroller has no incentive to deviate from $\boldsymbol{q}$ if

$$R_{pat}(\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n) \geq R_{pat}(\tilde{\boldsymbol{q}}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_n) \quad \forall \quad \tilde{\boldsymbol{q}} \in \Delta(\mathcal{A}_{pat})$$

and smuggler i has no incentive to deviate from $\boldsymbol{\pi}_i$ if

$$R_{smug(i)}(\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_i, \ldots, \boldsymbol{\pi}_n) \geq R_{smug(i)}(\boldsymbol{q}, \boldsymbol{\pi}_1, \ldots, \tilde{\boldsymbol{\pi}}_i, \ldots, \boldsymbol{\pi}_n) \quad \forall \quad \tilde{\boldsymbol{\pi}}_i \in \Delta(\mathcal{A}_{smug(i)}).$$

4.3.2 Selfish - Communication (S-C)

We now consider the case where the smugglers are still selfish, but communicate among themselves. We define a notion of equilibrium between the players in this case by using the concept of a 'correlated equilibrium', discussed in Aumann (1987). The concept developed by Aumann (1987) is that every player chooses a private action based on the observation of a publicly available stochastic signal, and if none have an incentive to deviate, then the signal is a correlated equilibrium. However, in our setting we do not wish the patroller to be able to observe the public signal, and instead we assume that they are only able to observe the distribution of it.

We define a signal as a random variable S which takes values in the joint action space of the smugglers, $\mathcal{A}_{smug} = \{0, 1\}^n$. Before the game starts, a joint action $\boldsymbol{a}$ is sampled from the signal S. Every smuggler observes the joint action $\boldsymbol{a}$, and smuggler i is directed (under action $\boldsymbol{a}$) to play the action a_i. Meanwhile, the patroller chooses a strategy $\boldsymbol{q}$ as in the previous case (S-NC). The patroller knows the distribution of the signal S, but does not observe the joint action $\boldsymbol{a}$ that has been sampled from it.

The patroller's strategy $\boldsymbol{q}$ and the smugglers' signal S form an equilibrium if and only if the patroller has no incentive to deviate to another strategy and no smuggler

has an incentive to deviate from the action suggested by the signal. For the patroller, this implies that $\boldsymbol{q}$ must satisfy

$$\mathbb{E}_{b\sim\boldsymbol{q},\boldsymbol{a}\sim S}\left[r_{pat}(b,\boldsymbol{a})\right] \geq \mathbb{E}_{b\sim\tilde{\boldsymbol{q}},\boldsymbol{a}\sim S}\left[r_{pat}(b,\boldsymbol{a})\right] \ \forall\, \tilde{\boldsymbol{q}} \in \Delta(\mathcal{A}_{pat}).$$

Smuggler i has no incentive to deviate if, for any joint action $\boldsymbol{a}$ shown by the signal S with non-zero probability ($\mathbb{P}(S=\boldsymbol{a})>0$), there is no alternative action that would strictly increase their expected reward. Since in our game smuggler i's actions are only 0 or 1 we only need to consider whether

$$\mathbb{E}_{b\sim\boldsymbol{q}}\left[r_{smug(i)}(b,a_1,\ldots,a_i,\ldots,a_n)\right] \geq \mathbb{E}_{b\sim\boldsymbol{q}}\left[r_{smug(i)}(b,a_1,\ldots,1-a_i,\ldots a_n)\right]$$

Communication allows selfish smugglers to observe the actions that other smugglers will take, whereas in the case of no communication (S-NC) they can only observe the distribution over the actions. It seems natural to suppose that giving selfish smugglers the ability to communicate leads to equilibrium strategies in which they obtain larger expected rewards. We make this argument precise, and furthermore prove it in Section 4.6.

4.3.3 Cooperative (CP)

The final case we consider is that of cooperative smugglers. When the smugglers are cooperative, we consider them as a group of smugglers represented by a single player in the game. Therefore, the reward to the group of smugglers given that they play an action $\boldsymbol{a}$ and the patroller plays an action b is given by the cumulative reward

$$r_{smug}(b,a_1,\ldots,a_n) = \sum_{k=1}^{n} r_{smug(k)}(b,a_1,\ldots,a_n) \tag{4.3.4}$$

We then assume that every smuggler receives an equal share of the total reward received. An equilibrium in the cooperative setting is a Nash equilibrium in the two player game between the group of smugglers and the patroller.

As in the previous cases (S-NC and S-C), the patroller's strategy $\boldsymbol{q}$ is a probability distribution over their actions $\mathcal{A}_{pat}$. The smugglers choose a strategy $\boldsymbol{\pi}$ over their joint action space $\mathcal{A}_{smug} = \bigtimes_{i=1}^{n} \mathcal{A}_{smug(i)}$. The probability of the smugglers taking a joint action $\boldsymbol{k}$ is given by $\boldsymbol{\pi}^{\boldsymbol{k}} = \mathbb{P}(\boldsymbol{a} = \boldsymbol{k})$. Therefore the patroller's expected reward, given that actions are chosen at random according to the strategies $\boldsymbol{q}$ and $\boldsymbol{\pi}$, is given by

$$R_{pat}(\boldsymbol{q}, \boldsymbol{\pi}) = \mathbb{E}_{b\sim\boldsymbol{q}, \boldsymbol{a}\sim\boldsymbol{\pi}} \left[r_{pat}(b, \boldsymbol{a}) \right]$$

and the smugglers' expected reward is given by,

$$R_{smug}(\boldsymbol{q}, \boldsymbol{\pi}) = \mathbb{E}_{b\sim\boldsymbol{q}, \boldsymbol{a}\sim\boldsymbol{\pi}} \left[r_{smug}(b, \boldsymbol{a}) \right].$$

The strategies $\boldsymbol{q}$ and $\boldsymbol{\pi}$ are a Nash equilibrium if and only if neither player has an incentive to deviate from their strategy. We defined the conditions for a Nash equilibrium for selfish smugglers in Section 4.3.1, and we have similar conditions in the case of cooperative smugglers. Formally, the patroller has no incentive to deviate from $\boldsymbol{q}$ if

$$R_{pat}(\boldsymbol{q}, \boldsymbol{\pi}) \geq R_{pat}(\tilde{\boldsymbol{q}}, \boldsymbol{\pi}) \quad \forall \quad \tilde{\boldsymbol{q}} \in \Delta(\mathcal{A}_{pat})$$

and the smugglers have no incentive to deviate from $\boldsymbol{\pi}$ if

$$R_{smug}(\boldsymbol{q}, \boldsymbol{\pi}) \geq R_{smug}(\boldsymbol{q}, \tilde{\boldsymbol{\pi}}) \quad \forall \quad \tilde{\boldsymbol{\pi}} \in \Delta(\mathcal{A}_{smug})$$

We have introduced the three different cases of smuggler behavior (S-NC, S-C and C) in our model. In Sections 4.5 and 4.6 we investigate equilibrium strategies in the respective cases.

4.4 Smuggler Marginal Probability Of Attacking

In this section we prove an important result concerning smuggler equilibrium strategies that holds in all of the cases of smuggler behavior that we consider. First, we introduce the notion of the 'marginal probability' with which a location is attacked. Denote the marginal probability of attack at location i by p_i, and note that p_i is the probability with which location i is attacked according to the strategy or signal chosen. An important feature of our model is that the patroller's expected reward depends only on the set of marginal probabilities, and not on the joint distribution of attacks across locations. Indeed, since the cost to the patroller due to undefended attacks, $g(x) = g_1 + g_2 x$, is a linear function of the number of successful attacks we have

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r_{pat}(b,\boldsymbol{a})\right] &= \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[ca_b - g(\alpha_b(\boldsymbol{a}))\right] \\ &= c\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[a_b\right] - \left(g_1 + g_2\sum_{i\neq b}\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[a_i\right]\right) \\ &= cp_b - \left(g_1 + g_2\sum_{i\neq b}p_i\right).\end{aligned} \tag{4.4.5}$$

We will use the vector $\boldsymbol{p} = (p_1, \ldots, p_n)$ to represent the set of marginal attack probabilities for locations $1, 2, \ldots, n$. We now prove that if two locations have different marginal attack probabilities, then in an equilibrium the location with the smaller attack probability must be protected with probability zero by the patroller.

Lemma 4.4.1. *There exists no equilibrium, with patroller strategy $\boldsymbol{q}$ and smuggler marginal attack probabilities $\boldsymbol{p}$, such that $p_j > p_i$ and $q_i > 0$ for some $i, j \in [n]$.*

Proof. Suppose that location i is protected with a probability greater than zero, $q_i > 0$. In an equilibrium, any action taken by the patroller with non-zero probability must be

in the set of best responses to the smugglers' actions. Therefore, defending location i must be a best response to the smugglers. However, we prove that the expected reward to the patroller for defending location j is strictly greater than they receive by defending location i. As in Eq.(4.4.5) we have,

$$\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}[r_{pat}(b,\boldsymbol{a})] = cp_b - \left(g_1 + g_2 \sum_{k\neq b} p_k\right).$$

We now compare the expected rewards to the patroller for protecting locations i and j.

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r_{pat}(i,\boldsymbol{a})\right] - \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r_{pat}(j,\boldsymbol{a})\right] &= \left[cp_i - \left(g_1 + g_2 \sum_{k\neq i} p_k\right)\right] - \left[cp_j - \left(g_1 + g_2 \sum_{k\neq j} p_k\right)\right] \\ &= (c+g_2)(p_i - p_j).\end{aligned}$$

It is an assumption of the model that both c and g_2 are positive, and an assumption of the lemma that $p_i < p_j$. Therefore, we have $\mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r_{pat}(i,\boldsymbol{a})\right] - \mathbb{E}_{\boldsymbol{a}\sim\boldsymbol{\pi}}\left[r_{pat}(j,\boldsymbol{a})\right] < 0$ and so the patroller has an incentive to deviate from taking the action of defending location i to defending location j. Thus, an equilibrium only exists if the patroller defends location i with probability zero, $q_i = 0$. □

Following on from Lemma 4.4.1 we prove that, in any equilibrium, if the patroller protects one location with probability strictly greater than another then the two locations must be attacked with equal marginal probability by the smugglers.

Lemma 4.4.2. *There exists no equilibrium, with patroller strategy $\boldsymbol{q}$ and smuggler marginal attack probabilities $\boldsymbol{p}$, such that $q_j > q_i$ and $p_i \neq p_j$ for some $i, j \in [n]$.*

Proof. First, suppose that $p_j < p_i$. By Lemma 4.4.1 if there exists an equilibrium we must have that location j is defended with probability zero, $q_j = 0$, contradicting the fact that $q_j > q_i$. Hence, we only need to consider $p_j \geq p_i$.

We now prove that we cannot have $p_j > p_i$ by assuming there exists such an equilibrium and finding a contradiction. Recall that since $p_j > p_i$, we must have by Lemma

4.4.1 that $q_i = 0$. We consider the different cases for smuggler behavior separately.

Firstly, we consider the case of selfish smugglers without communication (S-NC). Given that $p_i < 1$, the action of not attacking must be in the set of best responses for smuggler i to the patroller's strategy $\boldsymbol{q}$. If smuggler i chooses not to attack, their reward is zero. Thus, in order for $a_i = 0$ to be a best response, the expected reward for smuggler i attacking ($a_i = 1$) must be no greater than zero. Since $q_i = 0$ we can calculate the expected reward for the smuggler i attacking as

$$\mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\pi_1,\ldots,a_n\sim\pi_n}[r_{smug(i)}(b,a_1,\ldots,a_n)] = \sum_{k=1}^{n}\mathbb{P}(\alpha_b(\boldsymbol{a}) = k \text{ and } a_i = 1)f(k) \leq 0, \tag{4.4.6}$$

where (4.4.6) follows from (4.2.1) as a consequence of location i not being defended, implying that smuggler i will never be defended against. We reach a contradiction since the reward to the smuggler if $x \geq 1$ smugglers are successful, given by $f(x)$, is always strictly positive.

Secondly, we consider the case of selfish smugglers with communication (S-C). Since $p_j > p_i$ the signal S must suggest an action $\boldsymbol{a}$ such that $a_j = 1$ and $a_i = 0$ with non-zero probability. Define $x = \sum_k a_k$ to be the total number of attacks suggested when the signal displays the action $\boldsymbol{a}$ to the smugglers, and $q = \sum_k q_k a_k$ to be the probability that one of them is defended against. Conditional on the signal showing the action $\boldsymbol{a}$, the expected reward to the smuggler i for attacking is

$$(1-q)f(x+1) + qf(x),$$

since with probability $1-q$ there will be $x+1$ successful attacks and with probability q there will be x successful attacks. It was assumed that there is an equilibrium, which implies that smuggler i cannot have incentive to deviate away from not attacking ($a_i = 0$) to attacking ($a_i = 1$). Therefore, since the reward for not attacking is zero, we

require that,

$$(1-q)f(x+1)+qf(x)\leq 0.$$

We reach a contradiction since $f(x)>0$ and $f(x+1)>0$.

Finally, we consider the case of cooperative smugglers (CP). Due to the assumption that $p_j > p_i$, there must exist an action $\boldsymbol{a}$ taken with non-zero probability by the smugglers such that location j is attacked ($a_j = 1$) while location i is not attacked ($a_i = 0$). We compare the expected reward from taking action $\boldsymbol{a}$ with the expected reward from taking an alternative action $\tilde{\boldsymbol{a}}$, which we define to be identical to $\boldsymbol{a}$ except that we swap the attack at location j to location i. That is:

$$\tilde{a}_k = \begin{cases} a_j = 1 & \text{if } k = i, \\ a_i = 0 & \text{if } k = j, \\ a_k & \text{otherwise.} \end{cases}$$

Since action $\boldsymbol{a}$ is taken with non-zero probability, in order to have an equilibrium the action $\boldsymbol{a}$ must be a best response for the smugglers to the patroller's strategy $\boldsymbol{q}$. Therefore, the expected reward for choosing $\tilde{\boldsymbol{a}}$ cannot be strictly greater than the expected reward for choosing $\boldsymbol{a}$. It can be seen from (4.3.4) that when the patroller defends a location other than i or j, the reward to the smugglers is equivalent under actions $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$. Therefore we can simplify the difference in expected rewards to the smugglers for taking actions $\boldsymbol{a}$ and $\tilde{\boldsymbol{a}}$ by conditioning on the action of the patroller, as follows:

$$\begin{aligned} \mathbb{E}_{b\sim\boldsymbol{q}}\left[r_{smug}(b,\boldsymbol{a})\right] - \mathbb{E}_{b\sim\boldsymbol{q}}\left[r_{smug}(b,\tilde{\boldsymbol{a}})\right] &= \sum_k q_k r_{smug}(k,\boldsymbol{a}) - \sum_k q_k r_{smug}(k,\tilde{\boldsymbol{a}}) \\ &= \left[q_i r_{smug}(i,\boldsymbol{a}) + q_j r_{smug}(j,\boldsymbol{a})\right] - \left[q_i r_{smug}(i,\tilde{\boldsymbol{a}}) + q_j r_{smug}(j,\tilde{\boldsymbol{a}})\right] \\ &= q_j\left[r_{smug}(j,\boldsymbol{a}) - r_{smug}(j,\tilde{\boldsymbol{a}})\right], \end{aligned}$$

where the last line follows from the fact that $q_i = 0$. If we denote the number of attacks occurring at locations other than i and j by

$$x = \sum_{k \neq i,j} a_k$$

then, using the definition of the smugglers' reward function in (4.3.4), we have

$$\mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \boldsymbol{a})\right] - \mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \tilde{\boldsymbol{a}})\right] = q_j[(xf(x) - C) - (x+1)f(x+1)].$$

It is an assumption of the model that $(x+1)f(x+1) > xf(x) - C$. Therefore, we have that $\mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \boldsymbol{a})\right] - \mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \tilde{\boldsymbol{a}})\right] < 0$ and so the smugglers have an incentive to deviate from action $\boldsymbol{a}$ to action $\tilde{\boldsymbol{a}}$. This contradicts the assumption of an equilibrium, and therefore we cannot have $p_j > p_i$. □

Before stating our next result, we introduce some terminology to be used in the rest of the chapter. We say that the smugglers *attack uniformly* if every location is attacked with the same marginal probability; that is, $p_1 = p_2 = \cdots = p_n = p$. In the S-NC case, this implies that each smuggler attacks independently with probability $p \in [0, 1]$, whereas in the S-C and C cases, a certain number of smugglers $x \in [n]$ are randomly selected to attack (with each smuggler having the same probability of being selected). Similarly, we say that the patroller *defends uniformly* if every location is defended with the same probability; that is, $q_1 = q_2 = \cdots = q_n = 1/n$.

A consequence of Lemmas 4.4.1 and 4.4.2 is that if two locations are attacked with different marginal probabilities, then there cannot exist any patroller strategy $\boldsymbol{q}$ that results in an equilibrium. We state this as a theorem.

Theorem 4.4.3. *Under an equilibrium strategy, the smuggler marginal attack probabilities p_k must all be equal.*

Proof. Suppose, for a contradiction, that there exists an equilibrium in which the smug-

gler marginal attack probabilities $\boldsymbol{p}$ are not all equal. Hence, there exist locations i and j such that $p_i < p_j$. We denote the patroller's strategy in the equilibrium by $\boldsymbol{q}$. As a consequence of Lemma 4.4.1, the patroller must defend location i with zero probability, $q_i = 0$. There are then two cases: location j is either defended with non-zero probability ($q_j > 0$) or with zero probability ($q_j = 0$).

In the first case, if $q_j > 0 = q_i$ then as a consequence of Lemma 4.4.2 we have $p_j = p_i$, yielding an immediate contradiction.

In the second case, if $q_j = q_i = 0$ then there must exist another location k such that $q_k > 0$. Given that $q_k > q_i$, Lemma 4.4.2 implies that $p_k = p_i$. However, we then have $p_k = p_i < p_j$ and $q_k > q_j$ which contradicts the result of Lemma 4.4.2. □

Theorem 4.4.3 allows us to restrict the space of smuggler strategies that we consider to those with equal marginal attack probabilities when searching for equilibria.

4.5 Finding equilibria

In this section we consider the three different cases of smuggler behavior defined in Section 4.3 and detail in each of them how we can find equilibria.

4.5.1 Selfish - No Communication (S-NC)

We begin with the S-NC case. Having already proven that the marginal probabilities of attack at all locations must be equal in an equilibrium, we now focus on the patroller's strategy. We prove that the patroller must protect each location with the same probability, unless the smugglers attack with probability zero or one.

Proposition 4.5.1. *In the S-NC case, there exists no equilibrium such that the patroller defends non-uniformly and the smugglers each attack independently with probability* $p \in (0, 1)$.

Proof. Suppose, for a contradiction, that there exists an equilibrium of the type described in the proposition. Let i and j be two locations with $q_i \neq q_j$. We wish to compare the expected rewards that the smugglers at locations i and j obtain by attacking.

Given that there is an equilibrium where smugglers attack with probability $p \in (0,1)$, the actions of attacking and not attacking must both be in each smuggler's set of best responses to every other player's action. Consequently, since not attacking always gives a reward of zero, the expected reward for a smuggler attacking must also be zero. This implies that

$$\mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(i)}(b,a_1,\ldots,a_n)] = 0 = \mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(j)}(b,a_1,\ldots,a_n)] \tag{4.5.7}$$

We can give closed form expressions for both of the expectations in (4.5.7). Smuggler i attacks with probability $p \in (0,1)$. Conditional on them attacking, the patroller then defends smuggler i with probability q_i resulting in a cost of C. However, if location i is not defended, we know that one of the other locations must be defended. Therefore, if smuggler i attacks and is not defended against, then the number of successful attacks is given by the random variable $X_{n-2,p}+1$. Thus,

$$\mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(i)}(b,a_1,\ldots,a_n)] = p\left[(1-q_i)\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] - q_iC\right].$$

Following the same reasoning, we have that for smuggler j

$$\mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(j)}(b,a_1,\ldots,a_n)] = p\left[(1-q_j)\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] - q_jC\right].$$

Due to the equality in (4.5.7) we must have that

$$\begin{aligned}0 &= \mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(i)}(b,a_1,\ldots,a_n)] - \mathbb{E}_{b\sim\boldsymbol{q},a_1\sim\boldsymbol{\pi}_1,\ldots,a_n\sim\boldsymbol{\pi}_n}[r_{smug(j)}(b,a_1,\ldots,a_n)] \\ &= p\left[(1-q_i)\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] - q_iC\right] - p\left[(1-q_j)\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] - q_jC\right] \\ &= p(q_j-q_i)\left[\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] + C\right].\end{aligned} \tag{4.5.8}$$

However, we have assumed $p > 0$ and $q_j - q_i \neq 0$ in the statement of the proposition and $\mathbb{E}_{X_{n-2,p}}\left[f(X_{n-2,p}+1)\right] + C$ is positive due to the model assumptions. Therefore, the equality in (4.5.8) cannot hold and we reach a contradiction. □

In the S-NC case it will often be necessary to use binomially-distributed random variables to represent the number of attacking smugglers. To simplify the notation, we define $X_{n,p}$ as the binomial random variable with n trials and success probability p on each trial. From this point on, we consider only patroller strategies that defend uniformly. Whilst there exist other equilibria when the smugglers attack with probability zero or one, these edge cases are of less interest. Moreover, in the edge cases there also exist equilibria where the patroller defends uniformly, as we will later prove. We now turn our focus to the smugglers' strategies, and prove that there exists a unique probability of attack p^* that gives an equilibrium when the patroller defends uniformly.

We begin by calculating the expected reward to smuggler i for attacking with probability 1 when every other smuggler attacks with probability p. There are $n-1$ smugglers attacking independently with probability p, so the total number of these attacks is $X_{n-1,p}$. If $x \in \{0,\ldots,n-1\}$ other smugglers attack, as well as smuggler i, then the probability of no smuggler being caught is $[n-(x+1)]/n$, the probability of smuggler i being caught is $1/n$ and the probability of one of the other smugglers being caught is x/n. The payoffs to smuggler i in these cases are $f(x+1)$, $-C$ and $f(x)$ respectively.

Therefore, the expected reward to smuggler i is

$$w(p) := \mathbb{E}_{X_{n-1,p}}\left[\frac{n-(X_{n-1,p}+1)}{n}f(X_{n-1,p}+1)+\frac{X_{n-1,p}}{n}f(X_{n-1,p})-\frac{C}{n}\right] \quad (4.5.9)$$

In order to show that there exists a unique probability p^* giving an equilibrium, we begin by proving that (4.5.9) is strictly decreasing in p.

Lemma 4.5.2. *The expected reward to smuggler i for attacking with probability 1 while every other smuggler attacks with probability p, denoted by $w(p)$, is strictly decreasing with p.*

Proof. It is proved in Sah (1991) (appendices) that, for any function h, we have

$$\frac{\partial}{\partial p}\mathbb{E}_{X_{n-1,p}}[h(X_{n-1,p})] = (n-1)\sum_{k=0}^{n-2}\binom{n-2}{k}p^k(1-p)^{n-2-k}[h(k+1)-h(k)]. \quad (4.5.10)$$

We define the function h to be

$$h(k) = \frac{n-(k+1)}{n}f(k+1)+\frac{k}{n}f(k)-\frac{C}{n}.$$

Since f is a decreasing function, for any $k \in \{0,\dots,n-2\}$ we have

$$\begin{aligned}
& h(k+1)-h(k)\\
=\ & \left(\frac{n-(k+2)}{n}f(k+2)+\frac{k+1}{n}f(k+1)-\frac{C}{n}\right)-\left(\frac{n-(k+1)}{n}f(k+1)+\frac{k}{n}f(k)-\frac{C}{n}\right)\\
=\ & \frac{n-(k+2)}{n}[f(k+2)-f(k+1)]+\frac{k}{n}[f(k+1)-f(k)]\\
<\ & 0.
\end{aligned}$$

Hence, using (4.5.9) and (4.5.10), we have

$$\frac{\partial}{\partial p}w(p) = (n-1)\sum_{k=0}^{n-2}\binom{n-2}{k}p^k(1-p)^{n-2-k}[h(k+1)-h(k)] < 0$$

which completes the proof. □

Lemma 4.5.2 allows us to prove the existence and uniqueness of an equilibrium in the S-NC case.

Theorem 4.5.3. *In the S-NC case, an equilibrium exists only when each smuggler attacks independently with probability p^*, where p^* is uniquely specified in the interval* $[0, 1]$.

Proof. For a particular smuggler, the expected reward for attacking with probability 1 when all other smugglers attack with probability p is the function $w(p)$ defined in (4.5.9). On the other hand, the reward for not attacking is zero. As a result of Lemma 4.5.2 we have three possible cases: (4.5.9) can either be strictly negative for all $p \in [0, 1]$, strictly positive for all $p \in [0, 1]$ or strictly decreasing but equal to zero for some unique $p' \in [0, 1]$.

In the first case, where (4.5.9) is negative for all $p \in [0, 1]$, the best response of the smuggler is not to attack. Therefore, the only equilibrium must have $p = 0$.

Similarly, in the second case where (4.5.9) is positive for all $p \in [0, 1]$, the best response of the smuggler is to attack, and an equilibrium must have $p = 1$.

Finally, in the third case where $w(p') = 0$ for some $p' \in [0, 1]$. If $p \neq p'$ there is an incentive to deviate, as the unique best response must be to either attack or not attack (depending on the sign of $w(p)$). On the other hand, when $p = p'$ both actions give an expected reward of zero, and hence there is no incentive to deviate.

We have now proven in each of the cases that the smugglers have no incentive to deviate from their strategy. Now, we consider whether the patroller has an incentive to deviate. If smugglers attack uniformly, then the patroller's reward does not depend on which location they defend. Thus, they will not deviate from defending uniformly.

Consequently, there is a unique attack probability p^* such that the resultant strategy is a Nash equilibrium. This is either zero, one or $p' \in [0, 1]$, depending on which of the above cases applies. □

It follows from Theorem 4.5.3 that we have three possible types of equilibria in the S-NC case. Either $p^* = 0$, in which case no smuggler attacks, $p^* = 1$, in which case all smugglers attack and obtain positive expected rewards, or $p^* \in (0, 1)$, in which case the expected reward to every smuggler is zero.

4.5.2 Selfish - Communication (S-C)

We now consider the case where the smugglers are selfish but can communicate among themselves. Recall that in this case, smugglers observe a signal S that suggests an action to each of them. If no smuggler has an incentive to deviate from the signal's suggestion, then S forms an equilibrium with the patroller's strategy $\boldsymbol{q}$.

In the S-C case, there exist equilibria where the patroller does not defend uniformly. However, the analysis in this subsection proceeds under the assumption that the patroller defends uniformly in an equilibrium. In the other cases of smuggler behavior, if the smugglers play a mixed strategy in an equilibrium then the patroller must be defending uniformly. Therefore, by assuming the same property in the S-C case, we can provide comparisons between the different behavior cases (these can be found in Section 4.6). Furthermore, making this assumption allows for a more intuitive analysis of the smugglers' strategies, since we can exploit the symmetry of the model.

Theorem 4.4.3 states that the marginal attack probabilities p_i must all be equal in an equilibrium. In this section, we show that there exists a set $\mathcal{X}^* \subset \{0, \dots, n\}$ such that any stochastic signal that instructs $x \in \mathcal{X}^*$ smugglers to attack while resulting in equal marginal attack probabilities gives an equilibrium. Moreover, we show that there is no other signal that results in an equilibrium.

For some $\mathcal{X} \subset \{0, \dots, n\}$, define $\mathcal{S}(\mathcal{X})$ to be the set of signals that instruct $x \in \mathcal{X}$

smugglers to attack while also resulting in equal marginal attack probabilities. That is:

$$\mathcal{S}(\mathcal{X}) = \left\{ S \in \Delta(\{0,1\}^n) \;\middle|\; \begin{array}{l} p_i = p_j \;\forall\, i,j \in [n], \\ \sum_i a_i \in \mathcal{X} \text{ if } \mathbb{P}(S = \boldsymbol{a}) > 0 \end{array} \right\}.$$

The two constraints above respectively enforce that each location is attacked with the same marginal probability and that the number of smugglers attacking must be an element of $\mathcal{X}$.

Recall that we are assuming that the patroller defends uniformly. We denote the expected reward to a particular smuggler i for attacking, given that in total $x \in \{1, \ldots, n\}$ smugglers are attacking, by $u(x)$. If we have x smugglers attacking in total, then the probability that no smuggler is caught is $(n-x)/n$, the probability that smuggler i is caught is $1/n$ and the probability that another smuggler is caught is $(x-1)/n$. The payoffs to smuggler i in these cases are $f(x)$, $-C$ and $f(x-1)$ respectively. Thus, the expected reward to smuggler i conditional on them attacking is given by,

$$u(x) = \frac{n-x}{n} f(x) + \frac{x-1}{n} f(x-1) - \frac{C}{n}. \tag{4.5.11}$$

We first show that $u(x)$ is strictly decreasing with x.

Lemma 4.5.4. *The expected reward to each attacking smuggler, $u(x)$, is strictly decreasing with respect to the total number of attacking smugglers x.*

Proof. Suppose that we have $x \in \{1, \ldots, n-1\}$ smugglers attacking. If we were to add another attacking smuggler, then the change in expected rewards for the attacking smugglers would be

$$\begin{aligned} u(x+1) - u(x) &= \left[\frac{n-x-1}{n} f(x+1) + \frac{x}{n} f(x) - \frac{C}{n} \right] - \left[\frac{n-x}{n} f(x) + \frac{x-1}{n} f(x-1) - \frac{C}{n} \right] \\ &= \frac{n-x-1}{n} [f(x+1) - f(x)] + \frac{x-1}{n} [f(x) - f(x-1)] \end{aligned}$$

It is an assumption of the model that $f(x)$ is strictly decreasing. Furthermore, since we assumed $x \in \{1, \ldots, n-1\}$, we have $n - x - 1 \geq 0$ and $x - 1 \geq 0$. It follows from the above that

$$u(x+1) - u(x) < 0.$$

□

We now introduce the set $\mathcal{X}^*$, such that if $x^* \in \mathcal{X}^*$ smugglers are attacking, then none have any incentive to deviate. Formally, $\mathcal{X}^*$ is defined as follows:

$$\mathcal{X}^* = \begin{cases} \{0\} & \text{if } u(1) < 0, \\ \{n\} & \text{if } u(n) > 0, \\ \{x \in \mathbb{N} \mid u(x+1) \leq 0 \leq u(x)\} & \text{otherwise.} \end{cases} \tag{4.5.12}$$

From Lemma 4.5.4 it follows that exactly one of the cases in the definition of $\mathcal{X}^*$ must apply. We now prove that only the signals given by $S \in \mathcal{S}(x^*)$ result in equilibria in the S-C case.

Theorem 4.5.5. *Assume that the patroller defends uniformly. Then, in the S-C case, the signal S gives an equilibrium if and only if it is in the set $\mathcal{S}(\mathcal{X}^*)$.*

Proof. Suppose, for a contradiction, that there exists a signal $S \notin \mathcal{S}(\mathcal{X}^*)$ that results in an equilibrium. Recall that $\mathcal{S}(\mathcal{X}^*)$ consists of signals that satisfy the following constraints:

$$\sum_{\boldsymbol{a} \in \{0,1\}^n} (a_i - a_j)\mathbb{P}(S = \boldsymbol{a}) = 0 \; \forall \; i, j \in [n],$$

$$\sum_i a_i = x \in \mathcal{X}^* \text{ if } \mathbb{P}(S = \boldsymbol{a}) > 0.$$

Hence, given that $S \notin \mathcal{S}(\mathcal{X}^*)$, it must fail to satisfy at least one of these constraints. If the first constraint is not satisfied, then the smugglers do not all attack their respective

locations with equal marginal probabilities. However, this contradicts Theorem 4.4.3, so we have an immediate contradiction.

If the second constraint is not satisfied, then with positive probability an action $\boldsymbol{a}$ is taken such that $\sum_i a_i = x \notin \mathcal{X}^*$. By Lemma 4.5.4, the expected reward to attacking smugglers $u(x)$ is strictly decreasing in x. Hence, x is either smaller than every element of $\mathcal{X}^*$ or larger than every element of $\mathcal{X}^*$.

In the former case, where x is smaller than every element of $\mathcal{X}^*$, we have $u(x+1) > 0$ due to the definition of $\mathcal{X}^*$. Therefore, a smuggler who is not signalled to attack has an incentive to deviate from the signal and attack, so the signal cannot be an equilibrium. Similarly, in the case where x is larger than every element of $\mathcal{X}$, we have $u(x) < 0$ and hence a smuggler signalled to attack has an incentive to deviate by not attacking. This establishes that there cannot be a signal outside of $\mathcal{S}(\mathcal{X}^*)$ that gives an equilibrium.

We can also show that any signal $S \in \mathcal{S}(\mathcal{X}^*)$ must yield an equilibrium. The patroller has no incentive to deviate since each location is attacked with equal marginal probability, as a consequence of equation (4.4.5). Suppose the signal S prescribes the action $\boldsymbol{a}$ to the smugglers, where $\sum_i a_i = x \in \mathcal{X}^*$. An attacking smuggler, if there is one, receives an expected reward of $u(x) \geq 0$ and so has no incentive to deviate and stop attacking. Similarly, any non-attacking smuggler would receive an expected reward of $u(x+1) \leq 0$ by attacking and therefore has no incentive to deviate. □

4.5.3 Cooperation (CP)

Finally, we consider the case where the smugglers are cooperating. We aim to find the joint strategy for the smugglers that attacks all locations with equal probability, whilst offering no incentive to deviate.

As in previous cases, we begin by considering the patroller's strategy in an equilibrium. We first prove that if the patroller defends non-uniformly, then the smugglers must all take the same deterministic action (either 'attack' or 'do not attack'). An

important consequence of this result is that it is easy to characterize all equilibria that exist with the patroller defending non-uniformly, and we can then focus on other equilibria in which the smugglers have more interesting behavior.

Proposition 4.5.6. *If the patroller's strategy $\boldsymbol{q}$ does not defend uniformly, then the only possible actions in best response by the smugglers are either to all attack or to all not attack.*

Proof. Suppose that the patroller defends non-uniformly. Let i be a location defended with maximum probability, so $q_i = \max_k\{q_k\}$. Choose a different location j such that $q_i > q_j$, which must exist due to the assumption of uneven defending by the patroller. Consider an action $\boldsymbol{a}$ taken by the smugglers with non-zero probability in the equilibrium. We show that the action $\boldsymbol{a}$ cannot have $a_i = 1 > 0 = a_j$, otherwise the smugglers would have an incentive to deviate. We consider an action $\tilde{\boldsymbol{a}}$ that the smugglers could deviate to, where the attack from i to j is swapped whilst keeping every other attacking decision remains the same. That is:

$$\tilde{a}_k = \begin{cases} 0 & \text{if } k = i \\ 1 & \text{if } k = j \\ a_k & \text{otherwise.} \end{cases}$$

Switching from action $\boldsymbol{a}$ to $\tilde{\boldsymbol{a}}$ results in the cooperative smugglers strictly increasing their expected reward, since location i is defended with higher probability than location

j. Indeed, if $x = \sum_i a_i$ denotes the total number of attacks then we have

$$\begin{aligned}
&\mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \tilde{\boldsymbol{a}})\right] - \mathbb{E}_{b\sim \boldsymbol{q}}\left[r_{smug}(b, \boldsymbol{a})\right] \\
=\ & \sum_k q_k[r_{smug}(k, \tilde{\boldsymbol{a}} - r_{smug}(k, \boldsymbol{a})] \\
=\ & q_i\left[xf(x) - ((x-1)f(x-1) - C)\right] + q_j\left[(x-1)f(x-1) - C - xf(x)\right] \\
=\ & (q_i - q_j)\left[xf(x) - ((x-1)f(x-1) - C)\right] \\
>\ & 0
\end{aligned}$$

and hence there is an incentive to deviate. Therefore, if action $\boldsymbol{a}$ is chosen by the smugglers with non-zero probability in an equilibrium, we must have $a_i \leq a_j$. However, it cannot be the case that $a_i < a_j$, since in order to achieve equal marginal probabilities of attack (as required by Theorem 4.4.3) there would then need to be some action $\tilde{\boldsymbol{a}}$ taken with non-zero probability such that $\tilde{a}_i > \tilde{a}_j$. Therefore we must have $a_i = a_j$ in an equilibrium.

We have shown that $a_i = a_j$ for specific locations i and j. We now look across all the locations. The set of locations $[n]$ can be divided into two disjoint subsets, defined as

$$\mathcal{I} := \arg\max\{q_k\}$$

and

$$\mathcal{J} := [n] \setminus \mathcal{I}.$$

We can apply the previous argument to any $i \in \mathcal{I}$ and $j \in \mathcal{J}$. Therefore, every action $\boldsymbol{a}$ in the equilibrium must have $a_i = a_j$ for each $i \in \mathcal{I}$ and $j \in \mathcal{J}$. Consequently, the only possible actions for the smugglers in an equilibrium are $(0, \ldots, 0)$ and $(1, \ldots, 1)$, as required. □

As a consequence of Proposition 4.5.6, we can describe all equilibria in the case

of cooperating smugglers when the patroller does not defend uniformly. The patroller could choose any strategy, provided that every location is defended with non-zero probability. The smugglers' strategy must then be either to attack every location, attack no locations, or use a mixed strategy between the two if there is not a unique best response.

Having considered the case of the patroller defending non-uniformly, we now restrict our attention to equilibria where the patroller defends uniformly and aim to find the smugglers' best response to such a strategy. Recall from Section 4.5.2 that we define the set of random variables giving an equal marginal probability of attack, where only $x \in \mathcal{X}$ smugglers attack simultaneously, as $\mathcal{S}(\mathcal{X})$. We also define $u(x)$ as the expected reward to each attacking smuggler when x of them are attacking. It follows that the expected total reward to the smugglers is $xu(x)$. We now denote the set of values of x maximizing $xu(x)$ by $\mathcal{X}^*$. That is:

$$\mathcal{X}^* = \arg\max_x \{xu(x)\}$$

We can show that any strategy for the smugglers in $\mathcal{S}(\mathcal{X}^*)$ gives an equilibrium, assuming that the patroller defends uniformly. Moreover, the strategies in $\mathcal{S}(\mathcal{X}^*)$ are the only ones that can give an equilibrium, as any other strategy will give some player an incentive to deviate.

Theorem 4.5.7. *Assume that the patroller defends uniformly. Then a strategy $\boldsymbol{\pi}$ for the smugglers is an equilibrium if and only if $\boldsymbol{\pi} \in \mathcal{S}(\mathcal{X}^*)$.*

Proof. The proof of Theorem 4.5.7 follows very similar logic to that of Theorem 4.5.5. Suppose we have a strategy $\boldsymbol{\pi}$ that isn't included in $\mathcal{S}(\mathcal{X}^*)$. Then $\boldsymbol{\pi}$ must violate at

least one of the two constraints enforced by $\mathcal{S}(\mathcal{X}^*)$, which are:

$$\sum_{\boldsymbol{a}\in\{0,1\}^n} (a_i - a_j)\mathbb{P}(S = \boldsymbol{a}) = 0 \;\forall\; i, j \in [n],$$

$$\sum_i a_i = x \in \mathcal{X}^* \text{ if } \mathbb{P}(S = \boldsymbol{a}) > 0.$$

If the first constraint is not satisfied, then by Theorem 4.4.3 we cannot have an equilibrium since the locations are not attacked with equal marginal probabilities.

If the second constraint is not satisfied, then with non-zero probability there is an action $\boldsymbol{a}$ taken such that $\sum_k a_k \notin \mathcal{X}^*$. However, if the smugglers deviate from $\boldsymbol{a}$ to an action $\tilde{\boldsymbol{a}}$ such that $\sum_k \tilde{a}_k \in \mathcal{X}^*$, then their expected reward increases. Therefore, any strategy for the smugglers that isn't in the set $\mathcal{S}(\mathcal{X}^*)$ cannot be an equilibrium.

We now prove that every strategy $\boldsymbol{\pi} \in \mathcal{S}(\mathcal{X}^*)$ gives an equilibrium under the assumption of the patroller defending uniformly. The patroller has no incentive to deviate due to Theorem 4.4.3 since $\mathcal{S}(\mathcal{X}^*)$ enforces that each location is attacked with the same marginal probability. The smugglers have no incentive to deviate, since any number of attackers in $\mathcal{X}^*$ already maximizes their expected reward by definition. □

Our next result considers the special case where the total reward to the smugglers for making x successful attacks is decreasing with x. In this case we are able to give a more detailed result about the equilibria for cooperating smugglers. Specifically, we can prove that the number of attacking smugglers in an equilibrium cannot be greater than two.

Proposition 4.5.8. *Suppose $xf(x)$ is a decreasing function of x. Then the number of attacking smugglers in an equilibrium cannot be greater than two.*

Proof. Suppose the set $\mathcal{X}^*$ includes a value $x > 2$. We will prove that the smugglers would receive a strictly greater expected reward by making $x - 1$ attacks, thereby contradicting the definition of $\mathcal{X}^*$. When x smugglers attack the expected reward is

given by

$$xu(x) = x\left(\frac{n-x}{n}f(x) + \frac{x-1}{n}f(x^*-1) - \frac{C}{n}\right).$$

Comparing this with the expected reward when $x-1$ smugglers attack, we find that

$$\begin{aligned}
& xu(x) - (x-1)u(x-1) \\
= \; & x\left(\frac{n-x}{n}f(x) + \frac{x-1}{n}f(x-1) - \frac{C}{n}\right) - (x-1)\left(\frac{n-(x-1)}{n}f(x-1) + \frac{x-2}{n}f(x-2) - \frac{C}{n}\right) \\
= \; & x\frac{n-x}{n}f(x) + \left[x\frac{x-1}{n} - (x-1)\frac{n-(x-1)}{n}\right]f(x-1) - (x-1)\frac{x-2}{n}f(x-2) - \frac{C}{n} \\
= \; & \frac{n-x}{n}[xf(x) - (x-1)f(x-1)] + \frac{x-1}{n}[(x-1)f(x-1) - (x-2)f(x-2)] - \frac{C}{n}.
\end{aligned}$$

Given that $xf(x)$ is assumed to be decreasing with x and $C/n > 0$, we have $xu(x) - (x-1)u(x-1) < 0$. Therefore, the smugglers have an incentive to deviate by making $x-1$ attacks, contradicting the assumption that $x \in \mathcal{X}^*$. □

In this section we have analyzed the properties of equilibrium solutions in the S-NC, S-C and CP cases one-by-one,. The next section focuses on comparisons between the different cases.

4.6 Comparing Cases

In this section we investigate the similarities and differences between the equilibria found for the different cases of smuggler behavior in the previous section. Firstly, we show that there exists a cost of capture C such that the smugglers do not ever attack in an equilibrium.

Lemma 4.6.1. *Assume that the patroller defends uniformly. Then, if $C > (n-1)f(1)$, there cannot exist an equilibrium in which any smuggler attacks.*

Proof. If no smuggler attacks then in the selfish cases (S-NC and S-C) each smuggler has an expected reward of zero, and in the cooperative case (CP) the group of smugglers has an expected reward of zero. Recall that the reward function for an individual smuggler, given by (4.2.1), decreases with respect to the number of successful attacks made. Therefore, by showing that the expected reward for a single smuggler attacking is negative, we can show that there is no incentive to deviate.

Indeed, if the patroller defends uniformly, the expected reward for a single attacking smuggler is,

$$\frac{n-1}{n}f(1) - \frac{C}{n} < 0.$$

□

Secondly, we prove that increasing the cost of capture C or decreasing the rewards for smuggling items $f(x)$ results in fewer attacks when smugglers are selfish. In the non-communication case (S-NC) the expected number of attacks decreases, and in the communication case (S-C) the maximum number of attacks decreases.

Lemma 4.6.2. *Suppose $0 < C_1 < C_2$ and let f_1 and f_2 be decreasing functions with $f_1(x) > f_2(x)$ for all x. Then:*

1. *In the S-NC case, the expected number of attacks when the smuggler's cost of capture is C_1 is greater than when the cost of capture is C_2. In the S-C case, the maximum number of attacks at equilibrium when the smuggler's cost of capture is C_1 is greater than when the cost of capture is C_2.*

2. *In the S-NC case, the expected number of attacks when the smuggler reward function is f_1 is greater than when it is f_2. In the S-C case, the maximum number of attacks at equilibrium when the smuggler reward function is f_1 is greater than when it is f_2.*

Proof. Both statements within the lemma follow the same proof. We give the proof for the first statement (involving the cost of capture C) and note that the second statement is obtained using an analogous argument, replacing C_1 and C_2 by f_1 and f_2 as appropriate. We have:

$$w(p) = \mathbb{E}_{X_{n-1,p}}\left[\frac{n-(X_{n-1,p}+1)}{n} f(X_{n-1,p}+1) + \frac{X_{n-1,p}}{n} f(X_{n-1,p}) - \frac{C}{n}\right],$$
$$u(x) = \frac{n-x}{n} f(x) + \frac{x-1}{n} f(x-1) - \frac{C}{n}.$$

Denote the expected reward under C_1 by $w_1(p)$ or $u_1(x)$ (depending on which behavior case is being considered), and under C_2 by $w_2(p)$ or $u_2(x)$. We then have both $w_1(p) > w_2(p)$ for all $p \in [0,1]$ and $u_1(x) > u_2(x)$ for all $x \in \{1,\dots,n\}$.

In the S-NC case, suppose the smugglers attack with probability p_1^* in an equilibrium when the expected reward to each smuggler is given by w_1. We have that $w_2(p) < w_1(p)$, and in Lemma 4.5.2 it was shown that w_2 is strictly decreasing in p. There are three possibilities: either (i) there exists a p_1^* such that $w_1(p_1^*) = 0$, (ii) $p_1^* = 1$ with $w_1(1) > 0$ or (iii) $p_1^* = 0$ with $w_1(0) < 0$. In the first case, $w_2(p_1^*) < 0$ and $w_2(p)$ is strictly decreasing in p, implying $p_2^* \le p_1^*$. In the second case, it is impossible for the number of attacks to increase. In the third case we have $w_2(p) < w_1(p) < 0$ for all p and so $p_2^* = 0 = p_1^*$.

In the S-C case, assume that we have a set $\mathcal{X}_1^*$ that gives the set of all equilibria to be $\mathcal{S}(\mathcal{X}_1^*)$, under capture cost C_1. Define $\mathcal{X}_2^*$ similarly under C_2. Recall that $\mathcal{X}^*$ was defined as

$$\mathcal{X}^* = \begin{cases} \{0\} & \text{if } u(1) < 0, \\ \{n\} & \text{if } u(n) > 0, \\ \{x \in \mathbb{N} \mid u(x+1) \le 0 \le u(x)\} & \text{otherwise.} \end{cases}$$

In the first case with $\mathcal{X}_1^* = \{0\}$, we have $u_2(1) < u_1(1) < 0$ and thus $\mathcal{X}_2^* = \{0\}$. In the second case the number of attacks cannot increase. Consider the third case, and

choose an arbitrary $x_1^* \in \mathcal{X}_1^*$. Suppose we have some $x_2^* \in \mathcal{X}_2^*$ such that $x_2^* > x_1^*$. Then $u_1(x_1^* + 1) \geq u_1(x_2^*) > u_2(x_2^*) \geq 0$, which is a contradiction since by definition $u_1(x_1^* + 1) \leq 0$. □

Lemmas 4.6.1 and 4.6.2 both apply to multiple cases for the smuggler behavior, but still only consider one case at a time. The next result involves a comparison between the different behavior cases. Let us denote the values of the game to the smugglers as $v_{\text{S-NC}}$, $v_{\text{S-C}}$ and v_{CP} in the S-NC, S-C and CP cases respectively. Recall that in the cooperative case (CP) it is assumed that rewards are divided equally among the smugglers. Similarly, in the S-C and CP cases, we denote the numbers of attacking smugglers in an equilibrium as $\mathcal{X}_{\text{S-C}}^*$ and $\mathcal{X}_{\text{CP}}^*$ respectively.

We first show that, for any choice of parameters, the value to a smuggler in the S-NC case is less than the value in the S-C case, which in turn is less than the value in the CP case.

Theorem 4.6.3. *Given any set of parameters, the values $v_{S\text{-}NC}$, $v_{S\text{-}C}$ and v_{CP} satisfy*

$$v_{S\text{-}NC} \leq v_{S\text{-}C} \leq v_{CP}.$$

Proof. We begin with the first inequality. Recall from the S-NC case that, in an equilibrium, the smugglers attack uniformly and each smuggler receives an expected reward of zero unless they are attacking with probability one (the edge case), in which case they may receive positive expected rewards.

In the first case, where the expected reward is zero, the inequality $v_{\text{S-NC}} \leq v_{\text{S-C}}$ follows trivially from the fact that the expected reward to each smuggler in the S-C case must be non-negative.

In the second case, we can show that if every smuggler is attacking with probability one in the S-NC case then they must also all be attacking with probability one in

the S-C case. Recall that the expected rewards to attacking smugglers in the S-NC and S-C cases are given by the functions w and u in (4.5.9) and (4.5.11) respectively. Given that smugglers have positive expected rewards in the S-NC case, it follows that $w(1) = u(n) \geq 0$. If we have equality then $v_{\text{S-NC}}$ and $v_{\text{S-C}}$ are both zero. If the inequality $w(1) > 0$ is strict, then we must have $\mathcal{X}^* = \{n\}$, implying that every smuggler attacks deterministically in the S-C case. Therefore, $v_{\text{S-NC}}$ and $v_{\text{S-C}}$ are equal.

To prove the second inequality, $v_{\text{S-C}} \leq v_{\text{CP}}$, we note that the cooperative strategy maximizes the smugglers' expected total reward by definition and therefore also maximizes the expected reward for each smuggler, as profits are split equally. □

Additionally, we are able to prove that the number of attacks made by the smugglers is greater in the S-C case than in the CP case. We use $\mathcal{X}^*_{\text{S-C}}$ and $\mathcal{X}^*_{\text{CP}}$ to denote the sets of possible numbers of attacks (defined in (4.5.12)) in the S-C and CP cases respectively.

Theorem 4.6.4. *For any $x^*_{S\text{-}C} \in \mathcal{X}^*_{S\text{-}C}$ and any $x^*_{CP} \in \mathcal{X}^*_{CP}$, we have*

$$x^*_{S\text{-}C} \geq x^*_{CP}.$$

Proof. Suppose for a contradiction that $x^*_{\text{S-C}} < x^*_{\text{CP}}$. Recall that $\mathcal{X}^*_{\text{S-C}}$ and $\mathcal{X}^*_{\text{CP}}$ are defined as

$$\mathcal{X}^*_{\text{S-C}} = \begin{cases} \{0\} & \text{if } u(1) < 0, \\ \{n\} & \text{if } u(n) > 0, \\ \{x \in \mathbb{N} \mid u(x+1) \leq 0 \leq u(x)\} & \text{otherwise} \end{cases}$$

and

$$\mathcal{X}^*_{\text{CP}} = \arg\max_x \{xu(x)\}$$

respectively. We consider the three cases in the definition of $\mathcal{X}^*_{\text{S-C}}$. First, if $u(1) < 0$, we have $x^*_{\text{S-C}} = 0$ and $u(1) < 0$. However, $u(1) < 0$ implies that $xu(x) < 0$ for all $x \geq 1$,

since $u(x)$ is strictly decreasing with x by Lemma 4.5.4. Therefore $\mathcal{X}^*_{\text{CP}} = \{0\}$ and we reach a contradiction.

Next, if $u(n) > 0$, we have $x^*_{\text{S-C}} = n < x^*_{\text{CP}}$. However, we cannot have more than n smugglers attacking, so a contradiction is reached immediately.

Finally, in the remaining case we have $u(x^*_{\text{S-C}}) \geq 0 \geq u(x^*_{\text{S-C}} + 1)$. Therefore, since $x^*_{\text{CP}} > x^*_{\text{S-C}}$ (by assumption) and u is a decreasing function, we must have $u(x^*_{\text{CP}}) < 0$. This is not possible in an equilibrium since the group of smugglers could increase their expected total reward by making zero attacks. □

It is not possible to strengthen the result of Theorem 4.6.4 by including a statement about the S-NC case. We demonstrate this using a counter-example in Section 4.7.

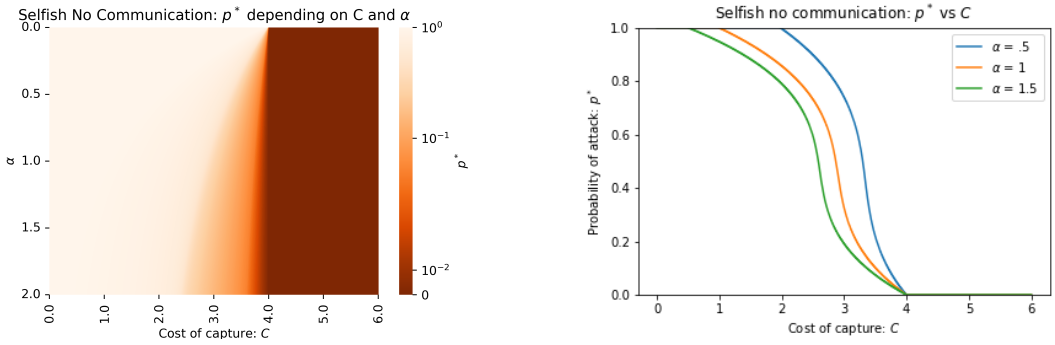
4.7 Examples

In this section we illustrate the results from earlier sections using a set of examples featuring a range of model parameters. We consider a border with $n = 5$ locations. The payoff to a single smuggler for being successful, given that a total of x smugglers are successful, is given by $f(x) = x^{-\alpha}$ where $\alpha > 0$ is a parameter to be varied. We note that when $\alpha > 1$ the function $xf(x) = x^{1-\alpha}$ is decreasing, meaning that the expected total reward for all smugglers decreases with the numbers of items successfully sent, meeting the assumptions of Proposition 4.5.8. However, if $\alpha < 1$ then the converse is true. We will also vary the cost of capture C from zero to $n + 1 = 6$ and show the effects on the smugglers' strategy in equilibria.

For the purposes of showing the strategies graphically, we adopt a convention that if the set $\mathcal{X}^*$ (used in the S-C and CP cases) includes multiple values then we select the smallest value.

4.7.1 Smuggler Strategy

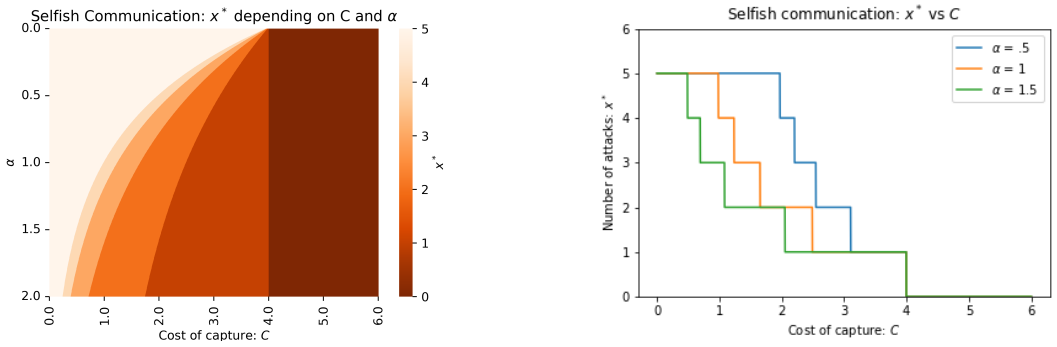
We begin by investigating how the smugglers attack in equilibria in each behavior case, illustrating the results from Section 4.5. Results based on comparisons between the different cases (based on Section 4.6) will be shown later. For the S-NC case, Figure 4.7.1 shows the probability of attack for each smuggler. Recall from Section 4.3.1 that there are three possible cases: (i) smugglers never attack, (ii) smugglers attack with probability $p \in (0, 1)$ and (iii) smugglers always attack. These cases are depicted in Figure 4.7.1(a) in brown, orange, and white respectively. Figure 4.7.1(b) plots the probability of attack against the cost of capture C for various α values. Next, for



(a) (b)

Figure 4.7.1: Smuggler strategy in the case of selfish smugglers without communication

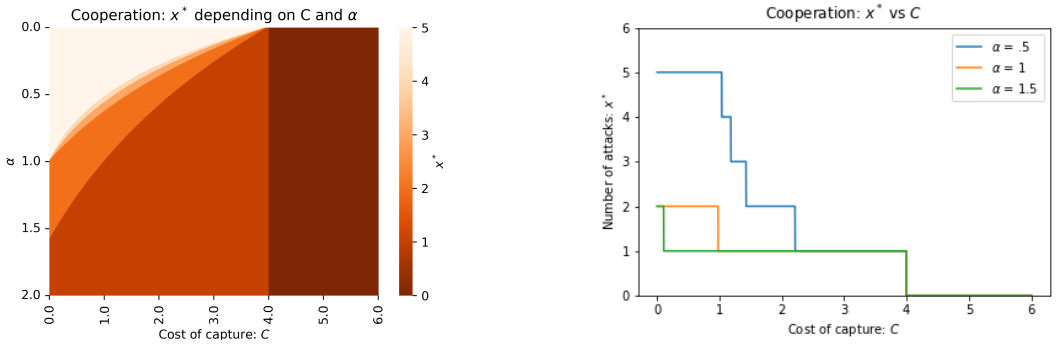
the S-C case, Figure 4.7.2 shows how the number of attacking smugglers x^* varies with C and α. In this case (unlike the S-NC case) the number of attacking smugglers is a deterministic function of the model parameters, and therefore Figure 4.7.2 has discrete regions, unlike Figure 4.7.1 which showed the probability of attack continuously changing as the parameters were varied. Figure 4.7.2(a) shows the number of attacking smugglers, defined using $\mathcal{X}^*$ in (4.5.12), and Figure 4.7.2(b) shows how x^* depends on C for various values of α. Finally, we consider the CP case. Like Figure 4.7.2 (for the S-C case), Figure 4.7.3 shows the number of attacking smugglers x^* as a deterministic function of the model parameters. Figure 4.7.3(a) shows the x^* values depend on C and α (recall that these are found as solutions of the maximization problem described



(a) (b)

Figure 4.7.2: Smuggler strategy in the case of selfish smugglers with communication

in Section 4.3.3). Figure 4.7.3(b) then shows how x^* depends on C for various values of α. Figure 4.7.3 also demonstrates that when $\alpha > 1$, since $xf(x)$ is decreasing, the number of attacks never exceeds two, as proved by Proposition 4.5.8. Figures 4.7.1-4.7.3

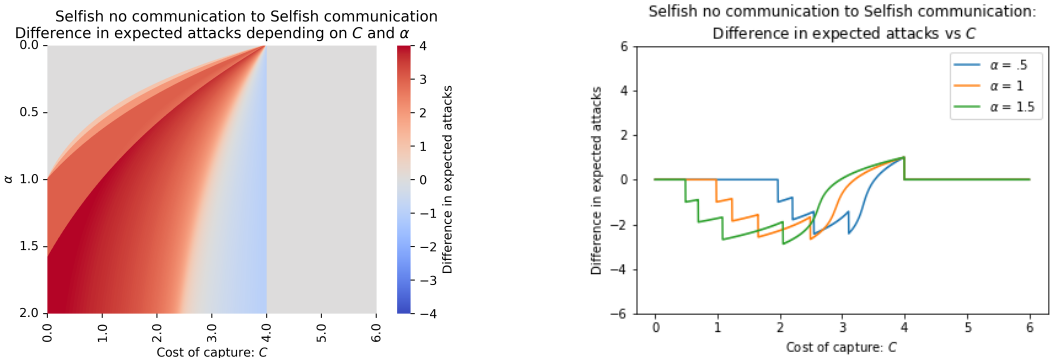


(a) (b)

Figure 4.7.3: Smuggler strategy in the case of cooperative smugglers

show some similarities between the different behavior cases. When the cost of capture C is at least $(n-1)f(1) = 4$, the smugglers never attack as proved by Lemma 4.6.1, which can be seen from the large regions in the lower parts of figures/ 4.7.1(a), 4.7.2(a) and 4.7.3(a). Additionally, the expected number of attacks decreases as C increases, or α increases (causing the function f to decrease), corroborating the result of Lemma 4.6.2. Looking at any horizontal or vertical slice in figures/ 1(a), 2(a) and 3(a) results in a decreasing number of attacks, shown explicitly in figures/ 4.7.1(b), 4.7.2(b) and 4.7.3(b).

Next, we compare the expected numbers of attacks across different behavior cases. Figure 4.7.4 illustrates the differences in expected numbers of attacks between the S-NC and S-C cases. Figure 4.7.4(a) shows that this number can either increase or decrease, depending on the parameters of the model. This confirms that the result of Theorem 4.6.4 cannot be extended to include a comparison between the S-NC and S-C cases. Figure 4.7.4(b) shows how the difference in the expected number of attacks varies with C and α. Figure 4.7.5 shows a similar comparison between the S-C and CP cases.



(a) (b)

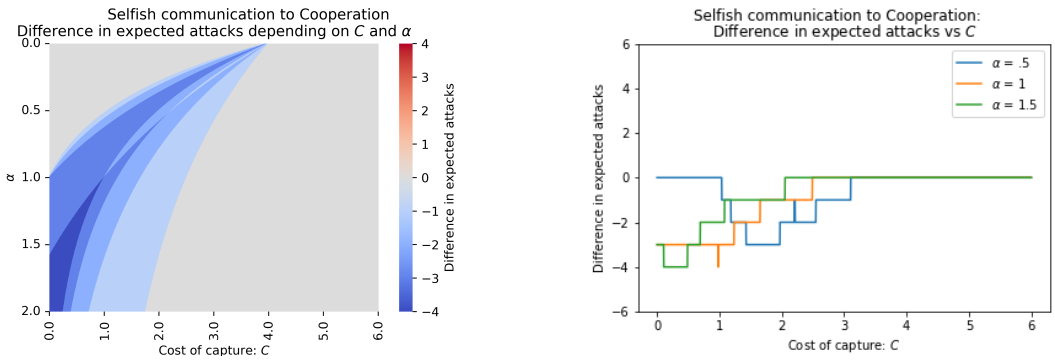
Figure 4.7.4: Differences in the expected number of attacks between the case of selfish smugglers without communication and the case of selfish smugglers with communication

Figure 4.7.5(a) shows that the expected number of attacks is lower in the CP case, corroborating the result of Theorem 4.6.4. Figure 4.7.5(b) shows how these differences depend on C and α.

4.7.2 Value of the game

Next, we discuss the value of the game to the smugglers in each of the different behavior cases. First we show how the value of the game depends on the model parameters, and then we compare the results for the different cases.

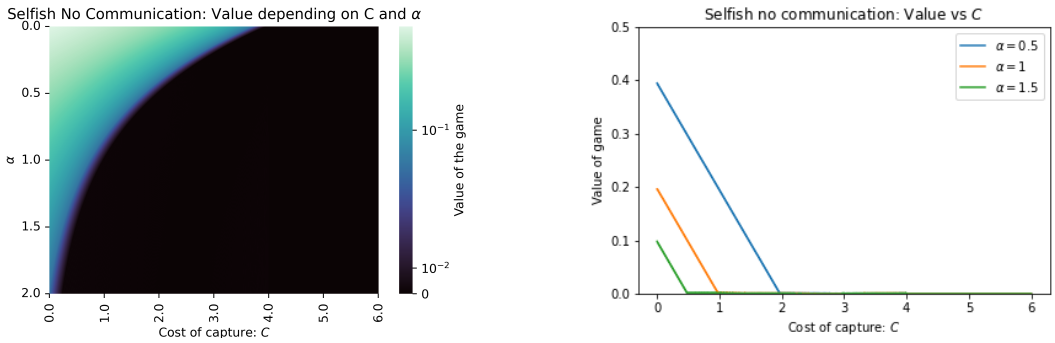
Figure 4.7.6 shows the expected values to the smugglers in the S-NC case. Figure 4.7.6(a) illustrates how these values depend on the model parameters. The main (dark) region of this figure shows that for many combinations of parameters, the expected



(a) (b)

Figure 4.7.5: Differences in the expected number of attacks between the case of selfish smugglers with communication and the case of cooperative smugglers

rewards to the smugglers are zero, as shown by the analysis in Section 4.3.1. The value of the game increases as either the cost of capture C decreases or the discount parameter α decreases. The top left part of Figure 4.7.6(a) shows the region in which every smuggler should always attack, as this results in a positive expected reward. Figure 4.7.6(b) shows how the value of the game depends on C for some fixed values of α. Next, Figure 4.7.7 shows the corresponding set of results in the S-C case. Figure



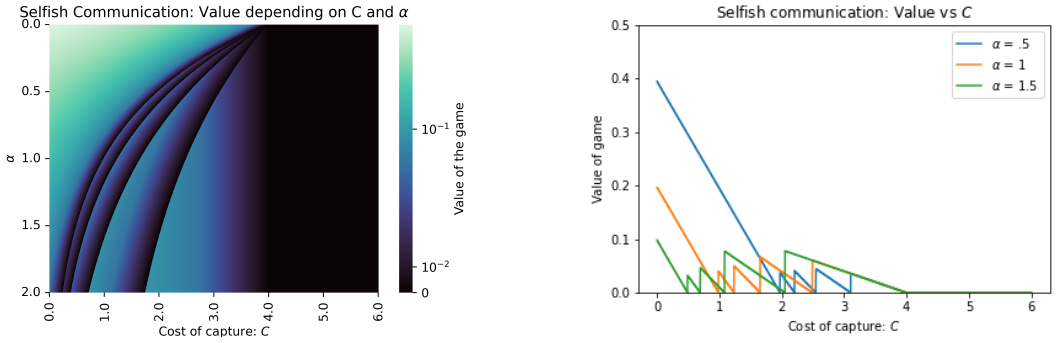
(a) (b)

Figure 4.7.6: Value of the game to selfish smugglers when there is no communication

4.7.7(a) shows how the expected reward for each smuggler depends on both C and α, and Figure 4.7.7(b) shows how the value depends on C for some fixed values of α.

Figure 4.7.7 may appear surprising at first, as (unlike in the S-NC and CP cases), the value of the game to the smugglers is non-monotonic in the model parameters.

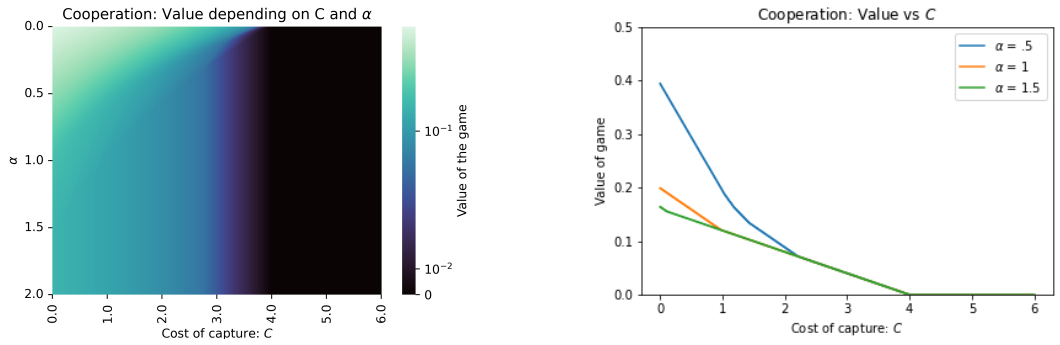
However, the non-monotonicity was proved by the analyses in Section 4.3.2. Suppose we have an equilibrium in which $x^* = x$ smugglers attack, for some $x \geq 1$. If we reduce the cost of capture C then the expected reward to each attacking smuggler $u(x)$ increases, and therefore so does the expected reward to each smuggler (unconditional on whether they attack). However, eventually the cost C decreases to a point where $u(x+1) = 0$, and the equilibrium is then for $x^* = x + 1$ smugglers to attack, resulting in expected rewards of zero. This explains the non-monotonic pattern shown in the figure. Finally, Figure 4.7.8 shows the expected values of the game in the CP case.



(a) (b)

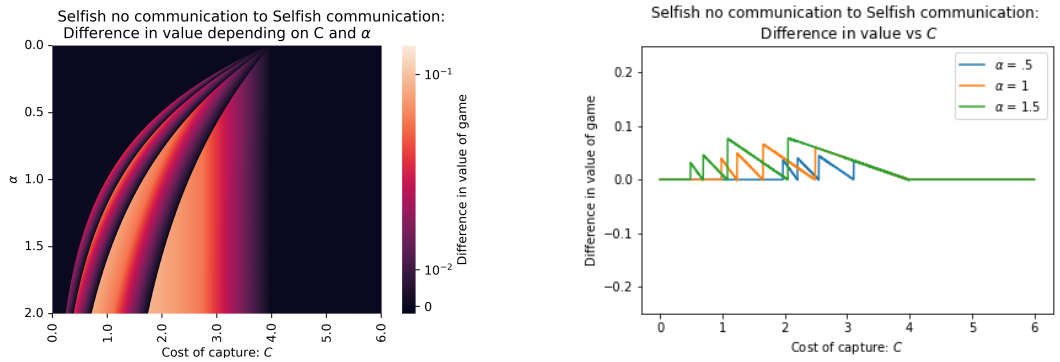
Figure 4.7.7: Value of the game to selfish smugglers when there is no communication

Figure 4.7.8(a) shows how these values depend on both C and α. Figure 4.7.8(b) shows how these values depend on C, for some fixed choices of α. We note that when only one smuggler is attacking, the value of α does not affect the reward received. Consequently, Figure 4.7.8(b) shows that the values for different choices of α become equal as C increases. In the remaining part of this section we compare the values of the game to the smugglers in the different behavior cases. Figure 4.7.9 shows the differences between expected rewards in the S-NC and S-C cases. Figure 4.7.9(a) illustrates these differences for various combinations of C and α, while Figure 4.7.9(b) shows how the differences depend on C for various fixed values of α. We observe that allowing the smugglers to communicate always improves the expected reward to each smuggler (as proven in Section 4.6). Figure 4.7.10 shows a similar comparison between the S-C and



(a) (b)

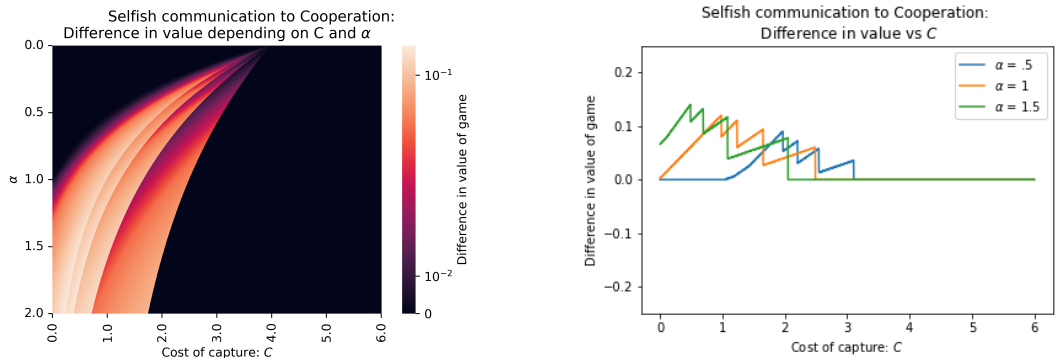
Figure 4.7.8: Values of the game to cooperative smugglers



(a) (b)

Figure 4.7.9: Difference in the value of the game between the case of selfish smugglers without communication and selfish smugglers with communication

CP cases. Figure 4.7.10(a) shows the difference in values for various combinations of C and α, while Figure 4.7.10(b) shows how the differences depend on C for some fixed value of α. By definition, the value of the game to the smugglers must improve in the CP case (as discussed in Section 4.6). Figure 4.7.10 indicates that this improvement is largest when the discount for sending more items (controlled by α) is large, resulting in a larger decrease in reward for sending multiple items. In the CP case, fewer attacks are made and therefore every successful attack results in a larger reward compared to the S-C case.



(a) (b)

Figure 4.7.10: Difference in the value of the game between selfish smugglers with communication and cooperative smugglers

4.8 Conclusion

In this chapter we have introduced a new model for patrolling a border against smugglers who receive diminishing marginal returns as more successful attacks are made. We have investigated equilibrium strategies under three different cases for the smuggler behavior: selfish smugglers without communication (S-NC), selfish smugglers with communication (S-C) and cooperative smugglers (CP). In each case we have analytically proven properties of the equilibrium strategies and also illustrated these using examples. We have found that the selfish smugglers benefit from being able to communicate on how many attacks will take place. Additionally, we have seen that being able to work together improves the smugglers' expected returns. Where possible, we have also proven results involving comparisons between the different cases.

Some interesting future directions would be to investigate how the results are affected when there are multiple patrollers, or when different types of items (with different associated rewards and penalties) are available for smuggling.

Chapter 5

A Stochastic Game for Border Patrol

In this chapter, we consider a different model to the one introduced in the previous chapter. There are similarities between the models, in both a single patroller defends a set of discrete locations that have a smuggler located there. However, there are many differences between the two models. Previously, we only considered a one-off game, whereas in this chapter we consider a stochastic game where there are an infinite number of decisions to be taken by either side. Furthermore, the payoffs in the game are different. We remove the dependence of the value of items on the total number of items smuggled in the previous chapter. However, we consider non-linear relationships between the penalty for capture and the quantity of items smuggled by the smugglers. Additionally, there is a movement cost for the patroller to relocate from one location to another. This chapter has been published as Darlington et al. (2023).

5.1 Introduction

Ranging from drug trafficking across the U.S.-Mexico border (Gutierrez and Henkel, 2021), to oil smuggling out of Nigeria (Ojewale, 2021), and illegal fishing in the con-

tinental shelf off South America (Goodman, 2021), the problem of how to patrol a border is fundamental to government organisations worldwide. How to patrol well is a challenging problem because it is infeasible to protect everywhere simultaneously due to constraints on resources, and thus a carefully thought out strategy is required. The associated trade-off is a complex one: if the patrols are too predictable the smugglers may be able to easily figure out where and when they can get through undetected. However, if the patrollers act too randomly they may not be adequately protecting the most vulnerable sections. In this work we introduce a stochastic game model for patrolling a border, detail how the strategies for both the patroller and smugglers can be found, and then analyse the solutions obtained.

Specifically, we consider a scenario where a single patroller attempts to stop a group of cooperating smugglers taking items across a border. The border here is thought of as being a finite set of locations which could be roads, border control posts or even an area of air, land or sea. The smugglers attempt to send some illicit items through these locations, and it is the patroller's goal to find an efficient strategy for stopping these items from getting through. It is assumed that there are known and fixed rewards and penalties that the smugglers receive or incur if they are respectively successful or not, proportional to the quantity of items they attempt to smuggle. Similarly, the patroller receives a reward or penalty depending on whether they stop the smugglers or not. There is a single smuggler fixed at each location. However, the patroller must traverse the geography of the border and pay a cost to do so. The patroller and smugglers make these decisions through time, needing to account for both their immediate reward and how their future rewards will be affected.

There is a significant operations research literature on patrol problems that focuses on modelling real-world situations. Examples include Sack and Urrutia, 1999 looking at the protection of galleries containing expensive paintings, or Richard, 1972 considering the daily patrol patterns of a police officer in the United States. An example where

developed methods have been implemented in practice concerns the protection of the Los Angeles International Airport and is by Pita et al., 2008. Pita et al.'s work models the problem as a Bayesian Stackelberg game to give the patrollers a randomised method to protect the airport from threats. The authors reported "very positive feedback about the deployment".

We consider a game theoretic approach to the problem of patrolling a border where the patroller and the smugglers take actions simultaneously. This was first considered by Alpern et al., 2011 who look at protecting against a single attempt by a smuggler that takes a fixed time to complete. An important assumption made in the chapter is that the outcome of the attempt results in a win or loss for the patroller. Lin et al., 2013 consider a similar model but introduce the possibility of the attempt taking a non-deterministic length of time to complete. Lin et al., 2014 further advance this work by considering the situation where the patroller has a chance to miss the attempt taking place. Another extension of this model is by McGrath and Lin, 2017 who solve a problem in which there is a non-trivial difference in both the time taken to travel around the locations and to check if an attempt is in progress at each location. The application of patrolling a border is considered by Papadaki et al., 2016 and Alpern et al., 2019, but it is still based on the assumption that the adversary makes only one attempt. Recent work in the patrolling literature includes Alpern et al., 2022b who consider a problem in which the patroller chooses whether or not to wear a uniform and Lin, 2022 who considers how to optimally patrol the perimeter of a location.

There are two key assumptions made by these papers that are not consistent with the problem of patrolling a border we consider. Firstly, in our setting a single successful smuggler is not catastrophic to the patroller. Instead, we have a small penalty incurred by the patroller that depends on the amount of items that are trafficked. Secondly, the normal-form game approaches discussed in previous papers can only consider trying to stop one attempt without taking into consideration what happens next. This means

that while we might catch the smugglers once, the patroller could be left exposed for an upcoming series of attempts. This motivates us to develop a new stochastic game model for patrolling a border. We will justify its benefits empirically with numerical experiments.

There are other bodies of work in the literature which look at a similar problem to ours. Grant et al., 2020 examine patrolling a border against opponents who make many small attempts to smuggle items. However, the smugglers are assumed to be acting at random whereas we make the stronger assumption that the smugglers pick actions strategically. The papers of Baston and Bostock, 1991 and Garnaev, 1994 discuss a stochastic game model for an inspection problem, applications of which include patrolling problems. However, their work differs from ours in that they consider a single smuggler, with a constraint on the number of times the patroller can attempt to capture the smuggler. Furthermore, the state in their stochastic game is the amount of time remaining in the game, whereas in our work the state of the game indicates the patroller's location. The model closest to ours is discussed in Filar and Schultz, 1986 and Filar, 1985, and is based on a different problem setting, involving a travelling inspector who checks factories to detect the illegal dumping of materials. An assumption of Filar's model is that the inspector's adversaries choose actions from a finite set. In contrast, the model presented in this chapter extends the action space to be a continuous set. Our model is able to more closely model situations in which items can be smuggled in vast or infinitely divisible quantities, and additionally we give a rigorous analysis of particular cases which would not be possible in the framework described by Filar. An infinite action space complicates the problem of finding Nash equilibria in the game. However, we detail how to overcome this challenge by using innovative solution algorithms.

The main contributions of the chapter are as follows:

- From the modelling perspective, we extend the smugglers' action set to a continuous interval. Having this larger set of alternatives gives a more realistic formula-

tion, especially so in the case in drug and oil smuggling scenarios, where the vast quantity of actions available to the smuggler intuitively forms a continuum.

- From the theoretical perspective, we provide an elucidation of the structure in the Nash equilibria of the game, yielding insight into the behavior of rational players in our model.

- From the methodological perspective, we develop new solution algorithms in order to overcome the increased computational challenge of finding Nash equilibria in our model. We prove that these algorithms find or converge to the optimal solution and, moreover, that they are computationally faster than existing methods in cases where a comparison is meaningful.

The rest of the chapter is organised as follows. In Section 5.2, we present our stochastic game framework for patrolling a border. Section 5.3 gives an overview of Nash equilibria that arise in the model, and establishes several of their properties. Section 5.4 provides an analysis of the methods to find Nash equilibria in the border patrol game. In Section 5.5, we make additional assumptions on the cost function which leads to more detailed characterisations of equilibria. An empirical analysis of the performance of our approach is given in Section 5.6, along with a discussion of the solutions to specific instances of our model. Section 5.7 concludes our chapter with a summary and suggestions for future work.

5.2 Model Description

We consider a border made up of n locations labelled from 1 to n inclusive. In the chapter we will use the notation $[n]$ to denote the set of all locations where $[n] = \{1, \ldots, n\}$. Time will be modelled in discrete steps $t = 0, 1, \ldots$. Such time steps are natural here, where decisions could be taken on an hourly or daily basis.

We present the model in this section by taking the smugglers collectively to be a single player. This will be the case throughout the chapter unless explicitly noted otherwise. Thus we look to define a stochastic game between two players: a single patroller and the smugglers. The patroller begins each time step t at some location s_t, which we take to be the current state of the system. Hence, the state space of the game is $\mathcal{S} = [n]$. The patroller picks a location b_t to defend, and the smugglers pick a quantity of items from the interval $[0, 1]$ to send to each location. We write the smugglers' action as $\boldsymbol{a}_t = (a_t^1, \dots a_t^n)$, where a_t^i is the quantity sent to the location i. Note that the assumption of actions in the unit interval is without loss of generality, since we can account for quantities from the interval $[0, q]$ for some $q > 0$ by a scaling of the actions. Hence, the action space of the patroller and smugglers respectively at each epoch are $\mathcal{A}_{pat} = [n]$ and $\mathcal{A}_{smug} = [0, 1]^n$. Both the patroller and the smugglers take an action simultaneously, with no knowledge of the action chosen by the opponent. The state of the system at the next time step is the previous action of the patroller, and so

$$\mathbb{P}(s_{t+1} = b \mid b_t = b, \boldsymbol{a}_t = \boldsymbol{a}, s_t = s) = \mathbb{P}(s_{t+1} = b \mid b_t = b) = 1 \tag{5.2.1}$$

for all $b \in \mathcal{A}_{pat}$, $\boldsymbol{a} \in [0, 1]^n$ and $s \in \mathcal{S}$. As we will see, the players can choose their actions according to some probability distribution which results in a random state transition in the game.

The patroller catches all items sent by the smugglers to the location they have chosen to defend. At every other location, the items are successfully smuggled. Smugglers receive a fixed reward of $r_i > 0$ for each unit of item smuggled through the location i. However, if caught, the smugglers must pay a penalty related to the amount smuggled. This is determined by the cost function $C : [0, 1] \to \mathbb{R}_+$. We assume that C is an increasing function with $C(0) = 0$. The patroller's payoff is equal to the negative of the smugglers' payoff, but she must additionally pay a cost for moving from one location to another. These movement costs are given by the parameters $m_{i,j} \geq 0$ $i, j \in [n]$. Thus,

the reward functions of the patroller and the smugglers respectively are as follows:

$$R_{pat}(b, \boldsymbol{a}, s) = -m_{s,b} + C(a_b) - \sum_{i \in [n] \setminus \{b\}} r_i a_i$$

$$R_{smug}(b, \boldsymbol{a}) = \sum_{i \in [n] \setminus \{b\}} r_i a_i - C(a_b).$$

The game continues for an infinite number of time steps, with rewards discounted at a rate of $\gamma \in [0, 1)$ for the patroller. Although smugglers can each have an individual discount rate of $\lambda_i \in [0, 1)$, we prove in Section 3 that the assumption that smugglers have a discount rate γ is without loss of generality.

A pure action is an action which a player is able to perform. In our case these are the elements of the sets $\mathcal{A}_{pat}$ and $\mathcal{A}_{smug}$ for the patroller and smugglers respectively. Instead of picking a pure action deterministically, players can draw an action according to a probability distribution over their pure actions, which may depend on the current state of the system s. A stationary mixed strategy for either player is a n-tuple of probability distributions over the pure actions of a player,

$$\boldsymbol{\Pi} = (\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^n) \in (\Delta([n]))^n$$

$$\boldsymbol{\Xi} = (\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^n) \in (\Delta([0, 1])^n)^n$$

where $\Delta(S)$ denotes the set of probability distributions with set S as their support and S^k is the k-ary Cartesian power of S for a natural number k. The results in subsequent sections will establish that it is sufficient to consider only stationary strategies, rather than strategies with a dependence on the current time step. The strategies for the patroller and smugglers respectively given that the state of the system is i are $\boldsymbol{\pi}^i$ and $\boldsymbol{\xi}^i$. Assuming the strategies are fixed over time, we write the expected discounted

reward for both players over an infinite horizon as

$$U_{pat}(\mathbf{\Pi}, \mathbf{\Xi}) = \mathbb{E}_{\mathbf{\Pi},\mathbf{\Xi},\mathbb{P}_0}\left[\sum_{t=0}^{\infty} \gamma^t R_{pat}(b_t, \boldsymbol{a}_t, s_t)\right],$$

and,

$$U_{smug}(\mathbf{\Pi}, \mathbf{\Xi}) = \mathbb{E}_{\mathbf{\Pi},\mathbf{\Xi},\mathbb{P}_0}\left[\sum_{t=0}^{\infty} \gamma^t R_{smug}(b_t, \boldsymbol{a}_t)\right]. \tag{5.2.2}$$

In (5.2.2) where expectations are taken with respect to the strategies of both players so that $b_t \sim \boldsymbol{\pi}^{s_t}$ and $\boldsymbol{a}_t \sim \boldsymbol{\xi}^{s_t}$, and also with respect to the probability distribution $\mathbb{P}_0$ over the initial state s_0. Since the outcome of one player depends on the action of the other, it is not possible to maximise the rewards of the players independently. We give the definition of a Nash equilibrium as first given by Nash, 1950.

Definition 5.2.1. *The strategies $\mathbf{\Pi}^*$ and $\mathbf{\Xi}^*$ for the patroller and smugglers respectively form a Nash equilibrium for the game if and only if,*

$$U_{pat}(\mathbf{\Pi}^*, \mathbf{\Xi}^*) \geq U_{pat}(\mathbf{\Pi}, \mathbf{\Xi}^*) \;\forall\; \mathbf{\Pi} \in (\Delta([n]))^n$$
$$U_{smug}(\mathbf{\Pi}^*, \mathbf{\Xi}^*) \geq U_{smug}(\mathbf{\Pi}^*, \mathbf{\Xi}) \;\forall\; \mathbf{\Xi} \in (\Delta([0,1]^n))^n.$$

Nash equilibria give the most natural solution for our model, in that they provide the best possible lower bound of the discounted expected reward to the patroller. This could be operationally important if, for example, the smugglers were to discover the strategy of the patroller and were able to optimise their strategy using this knowledge.

5.3 Properties of Nash Equilibria

We now seek to prove properties of the Nash equilibria in our model, which can help us to understand the behaviour of the patroller and the smugglers. Firstly, we note that the model described in the previous section falls into a class of stochastic games called

single controller stochastic games.

Definition 5.3.1. *Suppose we have an n-player stochastic game with players* $1, \ldots, n$*, with player* i *taking the action* a^i *from action set* $\mathcal{A}^i$*. Then the game is a single controller stochastic game with player* j *as the controller if and only if,*

$$\mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t^1 = a^1, \ldots, a_t^n = a^n) = \mathbb{P}(s_{t+1} = s' \mid s_t = s, a_t^j = a^j)$$

for all $s, s' \in \mathcal{S}$ *and* $a^i \in \mathcal{A}^i$ *for all players* i*. Filar and Vrieze, 2012*

It follows from Equation (5.2.1) that our model is a single controller stochastic game with the patroller as the controller. The single controller property leads to three results about Nash equilibria in our game: Lemma 5.3.2 proving that the smugglers can be assumed without loss of generality to be a single player, Proposition 5.3.3 showing discount rates of all players can be assumed to be equal without loss of generality, and Proposition 5.3.4 giving a zero-sum formulation of the game with equivalent Nash equilibria.

5.3.1 Aggregation of Smugglers

We first show that, without loss of generality, we can assume that the smugglers act as a single cooperating player. If the smugglers were acting independently of one another, then we would have an $n+1$ player game where each smuggler has a reward function,

$$R^i_{smug}(b, \boldsymbol{a}) = \begin{cases} r_i a_i, & b \neq i, \\ -C(a_i), & b = i, \end{cases}$$

for $i \in [n]$. The action space of smuggler i, $\mathcal{A}^i_{smug}$, is equal to the unit interval and his strategy, $\boldsymbol{\Xi}_i$, is in $\Delta([0,1])$. Therefore, a set of strategies for every player in this nonaggregated game is denoted $(\boldsymbol{\Pi}, \boldsymbol{\Xi}_1, \ldots, \boldsymbol{\Xi}_n)$. The patroller's reward function and action space remain the same, as do the state transitions and discount factors.

We can define a mapping from the reward function in the nonaggregated game to the aggregated game by,

$$\sum_{i=1}^{n} R^i_{smug}(b, \boldsymbol{a}) \to R_{smug}(b, \boldsymbol{a})$$

and a mapping from strategies in the nonaggregated game to the aggregated game by,

$$\left(\boldsymbol{\Pi}, \bigtimes_{i=1}^{n} \boldsymbol{\Xi}_i\right) \to (\boldsymbol{\Pi}, \boldsymbol{\Xi}).$$

Lemma 5.3.2. *Nash equilibria in the aggregated game coincide with those in the nonaggregated game in that if we have a Nash equilibrium in one game and map the strategies to the other, then it remains a Nash equilibrium.*

We omit the proof since a similar statement can be found in Filar, 1985, where an analytical proof is presented for their travelling inspector problem. The intuition behind the proof is that since the individual smugglers have independent reward functions and since their actions make no difference to the state transitions, neither combining nor splitting the smugglers create an incentive to deviate. A consequence of Lemma 5.3.2 is that we can choose whether to analyse the strategy for a single smuggler or the aggregated group, depending on which is more tractable in the context.

5.3.2 Discount Rates

We now move on to discuss the effect of players having different discount rates in our game for patrolling a border. We prove that if every player has an individual discount rate, then Nash equilibria are equivalent to those which occur when all players have the same discount rate as the patroller.

Proposition 5.3.3. *Suppose that* $(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}_1^*, \ldots, \boldsymbol{\Xi}_n^*)$ *is a Nash equilibrium for the nonaggregated game in which the patroller has a discount rate of* γ*, and smuggler* j *has a*

discount rate of $\lambda_j \in [0,1)$, $1 \leq j \leq n$. *Then* $(\mathbf{\Pi}^*, \mathbf{\Xi}_1^*, \ldots, \mathbf{\Xi}_n^*)$ *is a Nash equilibrium in every nonaggregated game in which the patroller has a discount rate of* γ, *and smuggler* j *has a discount rate of* $\tilde{\lambda_j} \in [0,1)$, $1 \leq j \leq n$.

Proof. Assume that $(\mathbf{\Pi}^*, \mathbf{\Xi}_1^*, \ldots, \mathbf{\Xi}_n^*)$ is a Nash equilibrium for the nonaggregated game in which the patroller has a discount rate of γ, and smuggler j has a discount rate of $\lambda_j \in [0,1)$, $1 \leq j \leq n$. We know that $(\mathbf{\Pi}^*, \mathbf{\Xi}^*)$ is a Nash equilibrium. We first show that after changing a single smuggler's discount rate we have an unchanged Nash equilibrium. Then, by induction we can apply this to every smuggler in turn to see that with discount rates $(\gamma, \tilde{\lambda}_1, \ldots, \tilde{\lambda}_n)$ we still have that $(\mathbf{\Pi}^*, \mathbf{\Xi}^*)$ is a Nash equilibrium.

Suppose that we alter smuggler j's discount rate to $\tilde{\lambda}_j$. First, we consider each player other than smuggler j. Since their discount factor is still the same, their expected reward is still the same, and thus they will not have any incentive to deviate from their strategy. The only player who may have an incentive to change their strategy is the smuggler j. However, since the patroller controls the state transitions in the game, the smuggler j can only maximise his instantaneous reward (further details can be seen in Corollary 1). A best response of smuggler j does not depend on their discount factor, and therefore they also do not have an incentive to deviate. □

The consequence of Proposition 5.3.3 is that we can assume without loss of generality that all players can be assumed to have a discount rate γ.

5.3.3 Zero-sum formulation of model

A two-player zero-sum stochastic game is defined as follows.

Definition 5.3.4. *A two-player stochastic game is zero-sum if the reward to one player is always equal to the negative of the reward to the other player.*

Whilst the model introduced in the previous section is not zero-sum, it only differs from one by the inclusion of the cost the patroller must pay to move around the

locations. We show that if the game is modified such that the smugglers are assumed to earn a reward equal to the cost of the movement of the patroller, then the Nash equilibria of the game are unchanged. The version of the game where the smugglers get this reward is clearly zero-sum.

Proposition 5.3.5. *Consider a stochastic game identical to the one introduced in the previous section, but where the reward function for the patroller and smugglers respectively are*

$$\tilde{R}_{pat}(b, \boldsymbol{a}, s) = R_{pat}(b, \boldsymbol{a}, s),$$
$$\tilde{R}_{smug}(b, \boldsymbol{a}, s) = -\tilde{R}_{pat}(b, \boldsymbol{a}, s) = R_{smug}(b, \boldsymbol{a}) + m_{s,b} = \sum_{i \in [n] \setminus \{b\}} r_i a_i + m_{s,b} - C(a_b).$$

The new game is a two player, zero-sum stochastic game. Furthermore, the Nash equilibria for the two games are identical.

Proof. Firstly, we consider whether the patroller has an incentive to deviate from their strategy $\boldsymbol{\Pi}^*$. The patroller's reward is the same in both games under any strategy taken by either player. Therefore, the patroller not having an incentive to deviate in one game implies that they have no incentive to deviate in the other.

Secondly, we explore whether the smuggler has any incentive to deviate from their strategy $\boldsymbol{\Xi}^*$. The patroller is the single controller in the stochastic game, and thus the smugglers can only try to maximize their instantaneous reward. The difference between the reward functions for the smugglers in the two games does not depend on their action $\boldsymbol{a}$ since,

$$R_{smug}(b, \boldsymbol{a}) - \tilde{R}_{smug}(b, \boldsymbol{a}, s) = m_{s,b}.$$

Therefore, they have no incentive to deviate from their strategy $\boldsymbol{\Xi}^*$ in one game if and only if they have no incentive to deviate from $\boldsymbol{\Xi}^*$ in the other.

Hence the Nash equilibria in the games coincide. □

Proposition 5.3.5 is similar to a result found in Hofbauer et al., 1998 for normal-form games; however, we were unable to find any such result in a stochastic game.

Please note that we shall consider this altered form of the game with reward functions $\tilde{R}$ for the remainder of the chapter. Proposition 5.3.5 is important since it allows us to apply a number of algorithms to find Nash equilibria which require that the game be zero-sum. Examples of these include algorithms for finite two-player zero-sum stochastic games developed by Shapley, 1953 and those developed for single controller games by Raghavan, 2003. A further consequence of Proposition 5.3.5 is that in our model there must indeed be a Nash equilibrium with stationary strategies. This follows from the result of Maitra and Parthasarathy, 1970, since in our game the reward to either player is continuous in the actions of both players and the state transition is deterministic.

In this section we have proven properties of the Nash equilibria in our stochastic game. However, the assumption of finite action spaces, made by Shapley, 1953 and Raghavan, 2003 does not hold in our model. The smugglers can take any action from the n-dimensional unit cube. Determining Nash equilibria remains a major challenge. This is the subject of Section 5.4.

5.4 Finding Nash Equilibria

There exist in the literature algorithms that can calculate Nash equilibria in two-player zero-sum stochastic games such as the one by Shapley, 1953. However, their assumption that the game is finite means that they are not directly applicable here. In this section, we present a method for determining Nash equilibria in our game. We begin by defining the value of a state s for the players.

Definition 5.4.1. *The value $\boldsymbol{V}_{pat}(s)$ of a state s to the patroller in the stochastic game is the expected reward to the patroller in a Nash equilibrium $(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*)$, given that the*

system starts in the state s, namely

$$\boldsymbol{V}_{pat}(s) = \mathbb{E}_{\boldsymbol{\Pi}^*,\boldsymbol{\Xi}^*}\left[\sum_{t=0}^{\infty}\gamma^t\tilde{R}_{pat}(b_t,\boldsymbol{a}_t,s_t)\;\middle|\;s_0=s\right].$$

The value of a state s for the smugglers, $\boldsymbol{V}_{smug}(s)$, is defined similarly.

The value of each state is unique and can be seen to solve the system of equations,

$$\boldsymbol{V}_{pat}(s) = \max_{\boldsymbol{\pi}\in\Delta([n])}\min_{\boldsymbol{a}\in[0,1]^n}\left[\sum_{b=1}^{n}\pi_b\left\{\tilde{R}_{pat}(b,\boldsymbol{a},s)+\gamma\boldsymbol{V}_{pat}(b)\right\}\right] \tag{5.4.3}$$

where π_b is the probability the patroller takes action b. This follows from Shapley, 1953 and Maitra and Parthasarathy, 1970. By (5.2.1), the transitions of system state are determined entirely by the patroller's choice of action. This is why in (5.4.3) we can deterministically know the system state resulting from any patroller action.

Shapley, 1953 proved that given any initial starting values $\{\boldsymbol{V}^0_{pat}(s)\mid s\in\mathcal{S}\}$ the sequence $\{\boldsymbol{V}^k_{pat}(s)\mid s\in\mathcal{S}\}_{k=1}^{\infty}$, determined by the recursion

$$\boldsymbol{V}^k_{pat}(s) = \max_{\boldsymbol{\pi}\in\Delta([n])}\min_{\boldsymbol{a}\in[0,1]^n}\left[\sum_{b=1}^{n}\pi_b\left\{\tilde{R}_{pat}(b,\boldsymbol{a},s)+\gamma\boldsymbol{V}^{k-1}_{pat}(b)\right\}\right]$$

converges to $\{\boldsymbol{V}_{pat}(s)\mid s\in\mathcal{S}\}$ as $k\to\infty$. When state and action spaces are finite, state values may be obtained by using linear programming to solve the maximisation problem within Shapley's iteration. However, since we assume that the action space of the smugglers is infinite, this approach is not available to us. Therefore, we look elsewhere to solve (5.4.3).

We begin by establishing properties about the smugglers' best response against any patroller strategy. If the smugglers take a best response against patroller strategy $\boldsymbol{\pi}=(\pi_1,\ldots,\pi_n)$ when the system state is s, the patroller receives a payoff which we

shall denote as $G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat})$

$$G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) = \min_{\boldsymbol{a}\in[0,1]^n} \left\{ \sum_{b=1}^{n} \pi_b [\tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma \boldsymbol{V}_{pat}(b)] \right\}.$$

If $\boldsymbol{V}_{pat}$ is the value function for the patroller, it will solve the following system of equations by (5.4.3).

$$\boldsymbol{V}_{pat}(s) = \max_{\boldsymbol{\pi}\in\Delta([n])} G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) \text{ for all } s \in [n].$$

Proposition 5.4.2. *The expected reward to the patroller for using strategy $\boldsymbol{\pi}$ in state s with value function $\boldsymbol{V}_{pat}$ when the smugglers play a best response can be calculated as:*

$$G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) = \sum_{b=1}^{n} \left[- \max_{a_b\in[0,1]} \{(1-\pi_b) r_b a_b - \pi_b C(a_b)\} + \pi_b(\gamma \boldsymbol{V}_{pat}(b) - m_{s,b}) \right] \quad (5.4.4)$$

Proof. The expected payoff to the patroller when the smugglers take a best response against them can be written as,

$$\min_{\boldsymbol{a}\in[0,1]^n} \left\{ \sum_{b=1}^{n} \pi_b [\tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma \boldsymbol{V}_{pat}(b)] \right\} = \min_{\boldsymbol{a}\in[0,1]^n} \left\{ \sum_{b=1}^{n} \pi_b \tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma \sum_{b=1}^{n} \pi_b \boldsymbol{V}_{pat}(b) \right\}. \quad (5.4.5)$$

The first sum can be rewritten if we expand upon the equation for the smugglers' reward function as follows,

$$\begin{aligned} \sum_{b=1}^{n} \pi_b \tilde{R}_{pat}(b, \boldsymbol{a}, s) &= \sum_{b=1}^{n} \pi_b \left[C(a_b) - \sum_{i\in[n]\setminus\{b\}} r_i a_i - m_{s,b} \right] \\ &= \sum_{b=1}^{n} [\pi_b C(a_b) + (\pi_b - 1) r_b a_b - \pi_b m_{s,b}]. \end{aligned}$$

Thus, if we consider this in Equation (5.4.5) we derive,

$$\begin{aligned}
&\min_{\boldsymbol{a}\in[0,1]^n}\left\{\sum_{b=1}^{n}\pi_b[\tilde{R}_{pat}(b,\boldsymbol{a},s)+\gamma \boldsymbol{V}_{pat}(b)]\right\}\\
=&\min_{\boldsymbol{a}\in[0,1]^n}\left\{\sum_{b=1}^{n}[\pi_b C(a_b)+(\pi_b-1)r_b a_b]-\pi_b m_{s,b}+\gamma\pi_b \boldsymbol{V}_{pat}(b)\right\}\\
=&\sum_{b=1}^{n}\left[\min_{a_b\in[0,1]}\{\pi_b C(a_b)+(\pi_b-1)r_b a_b\}-\pi_b m_{s,b}+\gamma\pi_b \boldsymbol{V}_{pat}(b)\right]\\
=&\sum_{b=1}^{n}\left[-\max_{a_b\in[0,1]}\{(1-\pi_b)r_b a_b-\pi_b C(a_b)\}+\pi_b(\gamma \boldsymbol{V}_{pat}(b)-m_{s,b})\right]
\end{aligned}$$

as required. □

The set of best responses for the smugglers against patroller strategy when the system state is s is given by,

$$\boldsymbol{a}(\boldsymbol{\pi},s)=\arg\min_{\boldsymbol{a}\in[0,1]^n}\left\{\sum_{b=1}^{n}\pi_b[\tilde{R}_{pat}(b,\boldsymbol{a},s)+\gamma \boldsymbol{V}_{pat}(b)]\right\}.$$

which we can simplify as a consequence of Proposition 5.4.2.

Corollary 5.4.3. *The set of best responses of the smugglers to the patroller's action $\boldsymbol{\pi}$ when the state of the game is s can be found as follows:*

$$\boldsymbol{a}(\boldsymbol{\pi},s)=\arg\min_{\boldsymbol{a}\in[0,1]^n}\left\{\sum_{b=1}^{n}\pi_b[\tilde{R}_{pat}(b,\boldsymbol{a},s)+\gamma \boldsymbol{V}_{pat}(b)]\right\}=(a_1(\pi_1,s),\ldots,a_n(\pi_n,s)).$$

where

$$a_i(\pi_i,s)=\arg\max_{a\in[0,1]}\{(1-\pi_i)r_i a-\pi_i C(a)\}$$

Proof. Follows from taking the argument of the minima in Proposition 5.4.2. □

From Corollary 5.4.3 we see the smugglers' best response to the patroller is a myopic one, and does not depend on the value function of either player, the discount rate γ or the system state s.

The function G is additively separable with respect to $\boldsymbol{\pi}$, and so we can write

$$G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) = \sum_{b=1}^{n} g_b(\pi_b, s, \boldsymbol{V}_{pat})$$

where

$$g_b(\pi_b, s, \boldsymbol{V}_{pat}) = -\max_{a\in[0,1]} \{(1-\pi_b) r_b a - \pi_b C(a)\} + \pi_b(\gamma \boldsymbol{V}_{pat}(b) - m_{s,b})$$

We now develop properties of the functions g_b, $b \in [n]$,which can be interpreted as the expected reward to the patroller for taking action b with probability π_b. These will be deployed to develop efficient approaches to the maximisation of G, and therefore the computation of Nash equilibria.

Lemma 5.4.4. *For every action available to the patroller $b \in [n]$, the expected reward to the patroller for taking that action $g_b(\cdot, s, \boldsymbol{V}_{pat}) : \mathbb{R} \to \mathbb{R}$ is concave and Lipschitz continuous with respect to the probability $\pi_b \in [0,1]$ that it is selected. Furthermore, the Lipschitz constant is $r_b + C(1) - (\gamma \boldsymbol{V}_{pat}(b) - m_{s,b})$ for a fixed system state s and value function $\boldsymbol{V}_{pat}$.*

Proof. We utilise Danskin, 1967 to establish the convexity of $\max_{a\in[0,1]} \{(1-\pi_b) r_b a - \pi_b C(a)\}$. The concavity of g_b in π_b, for a fixed s and $\boldsymbol{V}_{pat}$, is then immediate. Lipschitz continuity then follows since $[0,1]$ is compact. We now prove the Lipschitz constant.

Let $\delta > 0$, then the Lipschitz constant L can be taken as,

$$L \leq \max\left\{ \left| \frac{g_b(1+2\delta) - g_b(1+\delta)}{(1+2\delta) - (1+\delta)} \right|, \left| \frac{g_b(-2\delta) - g_b(-\delta)}{(-2\delta) - (-\delta)} \right| \right\}. \tag{5.4.6}$$

We can see that,

$$\arg\max_{a\in[0,1]}\{(1-\pi_b)r_b a-\pi_b C(a)\}=\begin{cases}1, \pi_b<0,\\ \\ 0, \pi_b>1.\end{cases}$$

$$\implies \max_{a\in[0,1]}\{(1-\pi_b)r_b a-\pi_b C(a)\}=\begin{cases}(1-\pi_b)r_b-\pi_b C(1), & \pi_b<0\\ \\ 0, & \pi_b>1\end{cases}$$

and so then we can evaluate (5.4.6) since,

$$\left|\frac{g_b(1+2\delta)-g_b(1+\delta)}{(1+2\delta)-(1+\delta)}\right|=\frac{1}{\delta}\left|\delta(\gamma \boldsymbol{V}_{pat}(b)-m_{s,b})\right|=\left|\gamma \boldsymbol{V}_{pat}(b)-m_{s,b}\right|$$

and,

$$\left|\frac{g_b(-2\delta)-g_b(-\delta)}{(-2\delta)-(-\delta)}\right|=\frac{1}{\delta}\left|-\delta C(1)-\delta r_b-\delta(\gamma \boldsymbol{V}_{pat}(b)-m_{s,b})\right|=\left|r_b+C(1)+\gamma \boldsymbol{V}_{pat}(b)-m_{s,b}\right|.$$

From the above calculations and (5.4.6) we conclude that,

$$\begin{aligned}L&=\max\left\{\left|r_b+C(1)+\gamma \boldsymbol{V}_{pat}(b)-m_{s,b}\right|,\left|\gamma \boldsymbol{V}_{pat}(b)-m_{s,b}\right|\right\}\\ &\leq r_b+C(1)-(\gamma \boldsymbol{V}_{pat}(b)-m_{s,b})\end{aligned}$$

since r_b and $C(1)$ are positive but $\gamma \boldsymbol{V}_{pat}(b)$ and $-m_{s,b}$ are negative. This concludes the proof. □

There is an existing literature to solve maximisation problems with an additively separable, concave objective function. Such problems are known as nonlinear knapsack or resource allocation problems. To approximate the continuous problem (5.4.7) we develop a scaled discrete problem (5.4.8). The scaling factor is denoted $K \in \mathbb{Z}$.

$$\begin{aligned} &\max \sum_{b=1}^{n} g_b(\pi_b, s, \boldsymbol{V}_{pat}) \\ &s.t. \sum_{b=1}^{n} \pi_b = 1 \\ &\pi_b \in [0,1] \end{aligned} \qquad (5.4.7) \qquad \begin{aligned} &\max \sum_{b=1}^{n} g_b(\pi_b, s, \boldsymbol{V}_{pat}) \\ &s.t. \sum_{b=1}^{n} \pi_b = 1 \\ &\pi_b \in \left\{ \frac{0}{K}, \frac{1}{K}, \ldots, \frac{K}{K} \right\} \end{aligned} \qquad (5.4.8)$$

We denote by $\boldsymbol{\pi}^*$ the optimal solution for the continuous problem (5.4.7) and by $\tilde{\boldsymbol{\pi}}_K$ the approximate solution obtained from the discrete scaled problem (5.4.8). Scaling by $K = n/\delta$ for some small $\delta > 0$ gives us the bound that $\|\boldsymbol{\pi}^* - \tilde{\boldsymbol{\pi}}_{n/\delta}\|_\infty \leq \delta$ by the proximity result of Hochbaum, 1994. Therefore, since for all b the function g_b is Lipschitz continuous we have that,

$$\begin{aligned} |G(\boldsymbol{\pi}^*, s, \boldsymbol{V}_{pat}) - G(\tilde{\boldsymbol{\pi}}_{n/\delta}, s, \boldsymbol{V}_{pat})| &\leq \sum_{b=1}^{n} |g_b(\pi_b^*, s, \boldsymbol{V}_{pat}) - g_b(\tilde{\pi}_{n/\delta,b}, s, \boldsymbol{V}_{pat})| \\ &\leq \sum_{b=1}^{n} [r_b + C(1) - (\gamma \boldsymbol{V}_{pat}(b) - m_{s,b})] |\pi_b^* - \tilde{\pi}_{n/\delta,b}| \\ &\leq \delta \sum_{b=1}^{n} [r_b + C(1) - (\gamma \boldsymbol{V}_{pat}(b) - m_{s,b})]. \end{aligned}$$

We conclude that $|G(\boldsymbol{\pi}^*, s, \boldsymbol{V}_{pat}) - G(\tilde{\boldsymbol{\pi}}_{n/\delta}, s, \boldsymbol{V}_{pat})| = \mathcal{O}(n\delta)$. The discrete resource allocation problem (5.4.8) can be solved greedily, as shown by Fox, 1966. This yields in Algorithm 4 for its solution.

Algorithm 4: Greedy Procedure by Fox, 1966

Initialise: $\tilde{\boldsymbol{\pi}}_K = (0, \ldots, 0), k = 0$

while $k < 1$ **do**

  Let,

$$j \in \arg\max_{b \in [n]} \left\{ g_b\left(\tilde{\pi}_b + \frac{1}{K}, s, \boldsymbol{V}_{pat}\right) - g_b(\tilde{\pi}_b, s, \boldsymbol{V}_{pat}) \right\}$$

  with ties decided by taking the lowest index.

  $\tilde{\pi}_{K,b} := \tilde{\pi}_{K,b} + \frac{1}{K}$ and $k := k + \frac{1}{K}$

end

Output: $\tilde{\boldsymbol{\pi}}_K$

While the complexity of Algorithm 4 is $\mathcal{O}(Kn) = \mathcal{O}(n^2/\delta)$ there exist more computationally efficient algorithms in the literature, such as Kaplan et al., 2019. This has a computational complexity of $\mathcal{O}(n \log K) = \mathcal{O}(n \log(n/\delta))$. In our examples, we consider both the algorithm by Fox, 1966 and by Kaplan et al., 2019. We have found that which is the quicker algorithm in practise can depend on the parameters of the problem.

We now leverage our ability to find a δ-optimal solution to the problem (5.4.7) in order to find the values of the states in the game via the iterative method of Shapley, 1953. This yields Algorithm 5.

Algorithm 5: Calculation of state values

Input: $\epsilon > 0$ and $\delta > 0$

Initialise: $\boldsymbol{V}^0_{pat}(s) = (0, \ldots, 0)$ and $k = 1$

1 **while** $\max_{s \in \mathcal{S}} \left\{ \left| \boldsymbol{V}^{k-1}_{pat}(s) - \boldsymbol{V}^{k}_{pat}(s) \right| \right\} > \epsilon$ **do**

2 **for** $s = 1, \ldots, n$ **do**

3 Find,

$$\begin{aligned} \boldsymbol{V}^k_{pat}(s) &:= \max_{\boldsymbol{\pi} \in \Delta([n])} \min_{\boldsymbol{a} \in [0,1]^n} \left[\sum_{b=1}^{n} \pi_b \left\{ \tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma \boldsymbol{V}^{k-1}_{pat}(b) \right\} \right] \\ &= \max_{\boldsymbol{\pi} \in \Delta([n])} G(\boldsymbol{\pi}, s, \boldsymbol{V}^{k-1}_{pat}) \end{aligned}$$

  using Algorithm 4 with $K = n/\delta$.

4 **end**

5 $k := k + 1$

6 **end**

Output: $\boldsymbol{V}^k_{pat}$

Once the state values have been calculated, the patroller's strategy $\boldsymbol{\Pi}^*$ can be identified as the value of $\tilde{\boldsymbol{\pi}}$ found in Step 3 of Algorithm 5. However, finding the smugglers' strategy $\boldsymbol{\Xi}^*$ which forms a Nash equilibrium with $\boldsymbol{\Pi}^*$ is a complex task without the addition of further assumptions on the parameters. In Section 5.5 we explore the characteristics of Nash equilibria under additional assumptions.

5.5 Behaviour of the Smugglers' Best Response

In this section, we focus on two different assumptions about the cost function C, which quantifies the losses of the smuggler when caught by the patroller. When the cost function is concave, we show that in a Nash equilibrium the smugglers only take actions in $\{0, 1\}$. This yields a more computationally efficient algorithm than Algorithm 5 in such cases, which is also guaranteed to find the optimal solution $\boldsymbol{\pi}^*$. When C is a strictly convex function, we show that the smugglers' strategy in equilibria takes actions deterministically.

5.5.1 Concave Cost Functions

We first examine the case in which the cost function C is a linear function, and then proceed to the case in which it is strictly concave. Under linearity, we prove that at least one of the actions zero or one lies within the set of best responses for each smuggler. Recall that we can calculate the set of best responses to patroller strategy $\boldsymbol{\pi}$ for the smuggler at a location b by

$$a_b(\pi_b, s) = \arg\max_{a \in [0,1]} \{(1 - \pi_b) r_b a - \pi_b C(a)\} .$$

Proposition 5.5.1. *If the cost function C is concave, then either 0 or 1 must lie within the set of best responses for the smuggler $a_b(\pi_b, s)$ to the patroller's strategy $\boldsymbol{\pi}$ at location b. Furthermore, if there exists an action $a \in (0, 1)$ which is a best response for the smuggler to the patroller's strategy $\boldsymbol{\pi}$ at location b, then C must be linear.*

Proof. The function $(1 - \pi_b) r_b a - \pi_b C(a)$ is convex in a, since C is concave. A maxima of a convex function on a convex set can always be found at an extreme points of that set, establishing the first result. If a maxima exists in the interior of the set, then the function must be constant on the set. In the case that $(1 - \pi_b) r_b a - \pi_b C(a)$ is constant, C must be linear. □

Proposition 5.5.1 allows us to simplify the game by reducing the action space of the smugglers.

Corollary 5.5.2. *If the cost function C is concave and $(\mathbf{\Pi}^*, \mathbf{\Xi}^*)$ is a Nash equilibrium in the border patrol game, then there exists a strategy $\tilde{\mathbf{\Xi}} \in (\Delta(\{0,1\}))^n$ such that $(\mathbf{\Pi}^*, \tilde{\mathbf{\Xi}})$ is a Nash equilibrium. We can therefore simplify the smugglers' strategy space from $(\Delta([0,1]))^n$ to $(\Delta(\{0,1\}))^n$.*

Proof. Suppose that the strategy $\mathbf{\Xi}^*$ takes an action $\boldsymbol{a}$ where $a_b \in (0,1)$ for some $b \in [n]$ with positive probability. Since $\mathbf{\Xi}^*$ must be a best response to $\mathbf{\Pi}^*$, Proposition 5.5.1 implies that C must be linear. The patroller's best response to $\mathbf{\Xi}^*$ when the system state is s gives a payoff of,

$$\max_{b \in [n]} \left\{ \mathbb{E}\left[\tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma V_{pat}(b) \right] \right\}$$

where the expectation is taken over $\boldsymbol{a} \sim \boldsymbol{\xi}_s$. However, when C is linear we have that

$$\begin{aligned} \mathbb{E}\left[\tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma V_{pat}(b) \right] &= \mathbb{E}\left[C(a_b) - \sum_{i \in [n] \setminus \{b\}} r_i a_i - m_{s,b} + \gamma V_{pat}(b) \right] \\ &= c\mathbb{E}[a_b] - \sum_{i \in [n] \setminus \{b\}} r_i \mathbb{E}[a_i] - m_{s,b} + \gamma V_{pat}(b) \end{aligned}$$

for some $c > 0$. Therefore, as if the smugglers instead take a strategy over the actions zero and one such that the expected quantity remains constant, then both players receive the same expected payoff. Hence, neither the patroller nor smuggler has incentive to deviate and so is a Nash equilbrium. This concludes the proof. □

From Corollary 5.5.2 we infer that the smuggler action space can be reduced to $\mathcal{A} = \{0,1\}^n$ without loss of generality. Having a finite action space for the smugglers means that the stochastic game is now finite and so Nash equilibria can be found using a linear programming formulation for single controller stochastic games. This

is as in Raghavan, 2003. Alternatively, we can use linear programming to maximise (6.3.3) in the iterative algorithm by Shapley, 1953. This is as in Filar and Vrieze, 2012. Corollary 2 also means that we can replace a strictly concave cost function with a linear cost function, provided that it takes the same values at the endpoints zero and one.

Corollary 5.5.3. *If the cost function C is concave then the Nash equilibria are equivalent to those in a game with identical parameters, but with cost function $\tilde{C}$ defined by $\tilde{C}(a) = C(1)a$.*

Proof. By Corollary 5.5.2, we have that the smuggler action space is $\{0,1\}^n$. Therefore, the cost function C is evaluated only at the points $a \in \{0,1\}$. Since $\tilde{C}(0) = C(0)$ and $\tilde{C}(1) = C(1)$, any Nash equilibria in the game with the cost function C must also be Nash equilibria in the game with the cost function $\tilde{C}$. □

We now look to simplify the expected reward to the patroller for playing a strategy $\boldsymbol{\pi}$ when the smugglers play a best response.

Lemma 5.5.4. *Assuming that the cost function C is concave, we can write the expected reward to the patroller for playing a strategy $\boldsymbol{\pi}$ when the smugglers play a best response as*

$$G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) = \sum_{b=1}^{n} \left\{ [\pi_b(C(1) + r_b) - r_b] \mathbb{1}\left(\pi_b \leq \frac{r_b}{C(1) + r_b}\right) + \pi_b(\gamma \boldsymbol{V}_{pat}(b) - m_{s,b}) \right\}. \tag{5.5.9}$$

Proof. Proposition 5.5.1 implies that $0 \in a_b(\pi_b, s)$ or $1 \in a_b(\pi_b, s)$. If we evaluate the smuggler's payoff at the two we get,

$$\begin{aligned} a = 0 &\implies (1 - \pi_b) r_b a - \pi_b C(a) = 0 \\ a = 1 &\implies (1 - \pi_b) r_b a - \pi_b C(a) = (1 - \pi_b) r_b - \pi_b C(1). \end{aligned}$$

This means that,

$$\mathbb{1}\left(\pi_b^s \leq \frac{r_b}{C(1)+r_b}\right) \in a_b(\pi_b, s)$$

and so,

$$\max_{a\in[0,1]}\{(1-\pi_b)r_b a - \pi_b C(a)\} = [(1-\pi_b)r_b - \pi_b C(1)]\mathbb{1}\left(\pi_b^s \leq \frac{r_b}{C(1)+r_b}\right)$$

Substituting this into the expression for G in Equation (5.4.4) gives the result in the statement of the lemma. □

A consequence of Lemma 5.5.4 is that we can now provide a computationally efficient method in Algorithm 6 to find the optimal value of G when the cost function C is concave.

Algorithm 6: Concave Cost Greedy maximization of G

Initialise: $\hat{\boldsymbol{\pi}} = (0, \ldots, 0)$

while $\sum_{b=1}^n \pi_b < 1$ **do**

  Define for all b,

$$x_b = \begin{cases} \dfrac{r_b}{C(1)+r_b}, & \hat{\pi}_b = 0, \\ 1 - \dfrac{r_b}{C(1)+r_b}, & \text{otherwise.} \end{cases}$$

  Choose arbitrarily,

$$j \in \arg\max_{b\in[n]} \left\{ \frac{g_b(\hat{\pi}_b + x_b) - g_b(\hat{\pi}_b)}{x_b} \right\}$$

  with ties decided by taking the lowest index.

  if $\sum_{b=1}^n \hat{\pi}_b + x_j \leq 1$ **then**

    Let $\hat{\pi}^j := \hat{\pi}^j + x_j$.

  else

    Let $\hat{\pi}^j := \hat{\pi}^j + (1 - \sum_{b=1}^n \hat{\pi}_b)$.

  end

end

Output: $\hat{\boldsymbol{\pi}}$

Theorem 5.5.5. *If the cost function C is linear, then the strategy $\hat{\boldsymbol{\pi}}$ given by Algorithm*

6 maximises the patroller's expected reward G given that the smugglers play a best response.

Proof. Recall that $\boldsymbol{\pi}^*$ is the optimal solution, $\tilde{\boldsymbol{\pi}}$ is found using Algorithm 4 and $\hat{\boldsymbol{\pi}}$ is constructed using Algorithm 6. Denote the value of $\hat{\boldsymbol{\pi}}$ after iteration l of Algorithm 6 by $\hat{\boldsymbol{\pi}}(l)$. We let,

$$K_m = m \prod_{b=1}^{n} [C(1) + r_b],$$

and denote the output of Algorithm 4 when using $K = K_m$ as $\tilde{\boldsymbol{\pi}}_{K_m}$. Furthermore, we respresent its value after a step of Algorithm 4 with value k by $\tilde{\boldsymbol{\pi}}_{K_m}(k)$. Since K_t is divisible by every $C(1) + r_b$, it follows that for every l there exists a k such that,

$$\sum_{b=1}^{n} \hat{\pi}_b(l) = \sum_{b=1}^{n} \tilde{\pi}_{K_m,b}(k)$$

and we denote this k by k_l.

We prove by induction that for every l we have $\hat{\boldsymbol{\pi}}(l) = \tilde{\boldsymbol{\pi}}_{K_m}(k_l)$. This is clearly true when $l = 0$ since $\hat{\boldsymbol{\pi}}(0) = \tilde{\boldsymbol{\pi}}_{K_m}(k_0) = \mathbf{0}$. Assume that for a given l we have $\hat{\boldsymbol{\pi}}(l) = \tilde{\boldsymbol{\pi}}_{K_m}(k_l)$. In Step 3 of Algorithm 6 we find,

$$j \in \arg\max_{b \in [n]} \left\{ \frac{g_b(\hat{\pi}_b(l) + x_b) - g_b(\hat{\pi}_b(l))}{x_b} \right\}.$$

Since g_b is linear on the interval $\pi_b \in [\hat{\pi}_b(l), \hat{\pi}_b(l) + x_b - \frac{1}{K}]$, this implies that,

$$j \in \arg\max_{b \in [n]} \left\{ g_b\left(\pi_b + \frac{1}{K}\right) - g_b(\pi_b) \right\}$$

for $\pi_b \in [\hat{\pi}_b(l), \hat{\pi}_b(l) + x_b - \frac{1}{K}]$. Thus, at every step between k_l and k_{l+1} in Algorithm 4 we increase $\tilde{\pi}^j$ and so,

$$\tilde{\pi}_{K_m,b}(k_{l+1}) = \tilde{\pi}_{K_m,b}(k_l) + \mathbb{1}(b = j)x_b = \hat{\pi}_b(k_l) + \mathbb{1}(b = j)x_b = \hat{\pi}_b(l + 1).$$

Thus, we have $\tilde{\boldsymbol{\pi}}_{K_m}(k_{l+1}) = \hat{\boldsymbol{\pi}}(l+1)$, so by induction we deduce that $\tilde{\boldsymbol{\pi}}_{K_m} = \hat{\boldsymbol{\pi}}$. For each $m \in \mathbb{N}$ the proximity result by Hochbaum, 1994 implies that,

$$\|\boldsymbol{\pi}^* - \tilde{\boldsymbol{\pi}}_{K_m}\|_\infty \leq \frac{n}{K_m}$$

which therefore means that since $K_m \to \infty$ as $m \to \infty$ we must have $\|\boldsymbol{\pi}^* - \tilde{\boldsymbol{\pi}}_{K_m}\|_\infty \to 0$. Since we know that $\|\boldsymbol{\pi}^* - \tilde{\boldsymbol{\pi}}_{K_m}\|_\infty = \|\boldsymbol{\pi}^* - \hat{\boldsymbol{\pi}}\|_\infty$ for all $m \in \mathbb{N}$ we must have that $\boldsymbol{\pi}^* = \hat{\boldsymbol{\pi}}$. □

The complexity of Algorithm 6 is only $\mathcal{O}(n)$, since the maximum number of iterations needed to complete is $n+1$ and each iteration has complexity $\mathcal{O}(1)$. We can see that it takes at most $n+1$ iterations, since once the probability of an action is increased twice, the algorithm must terminate.

So far, the discussion has focused only on determining a strategy $\boldsymbol{\Pi}^*$ for the patroller. We now consider how to find a strategy for the smugglers $\boldsymbol{\Xi}^*$ such that $(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*)$ is a Nash equilibrium in our model. Once we have found the value function $\boldsymbol{V}_{smug} = -\boldsymbol{V}_{pat}$, finding the smugglers' strategy can be found by taking:

$$\boldsymbol{\xi}^s \in \underset{\boldsymbol{\xi}^s \in \Delta(\{0,1\}}{\arg\max} \min_{b \in [n]} \mathbb{E}\left[\tilde{R}_{smug}(b, \boldsymbol{a}, s) + \gamma V_{smug}(b)\right]$$

for each $s \in [n]$. A linear program can efficiently solve this as in Filar and Vrieze, 2012.

5.5.2 Strictly Convex Cost Function

We now proceed to the case in which the cost function C is strictly convex with respect to the action taken by the smugglers. Algorithm 4 can give us an approximation for $\boldsymbol{\Pi}^*$, but as in the previous subsection, we still need to consider how we will calculate the smugglers' strategy $\boldsymbol{\Xi}^*$. The following lemma shows that under an assumption of strict convexity there can only be one choice, and additionally it is simple to find.

Lemma 5.5.6. *If the cost function C is strictly convex, then for any given patroller strategy $\boldsymbol{\pi}$, the smugglers have a single best response.*

Proof of Lemma 5.5.6

Proof. Recall that when the system state is s, the set of best responses for the smuggler at the location b to the patroller's strategy $\boldsymbol{\pi}$ is given by,

$$a_b(\pi_b, s) = \max_{a\in[0,1]} \{(1-\pi_b)r_b a - C(a)\}. \tag{5.5.10}$$

If C is strictly convex, then the function to be maximised in (5.5.10) is strictly concave in a_b. A strictly concave function can only have a single maximum in the interval $[0, 1]$, and therefore there can only be a single unique best response for the smuggler at b for any given patroller strategy. Applying this reasoning to each system state and every location, we can see that there must be a single strategy for the smugglers which is uniquely the best response to $\boldsymbol{\pi}$. □

From Lemma 5.5.6, we can quickly compute a best smuggler response $\boldsymbol{\Xi}$ to the patroller's strategy $\boldsymbol{\Pi}^*$. Since there exists a best response to $\boldsymbol{\Pi}^*$, and since there must exist at least one Nash equilibrium, then $(\boldsymbol{\Pi}^*, \boldsymbol{\Xi})$ must indeed be a Nash equilibrium.

5.6 Examples

In this section, we introduce three different examples and discuss how the analysis from previous sections helps to find Nash equilibria and how to understand them. We then go on to justify the use of a stochastic game model in terms of its benefits for the border patrol problem compared to the use of alternative models.

5.6.1 Example 1: Linear Border With Linear Cost Function

We begin by considering an example with a linear cost function. We compare the time taken to find Nash equilibria using the methods discussed in this chapter with existing methods in the literature. We can apply the latter, since by Corollary 5.5.2 we know that there exists a Nash equilibrium in which the smugglers' actions are supported by $\{0, 1\}^n$.

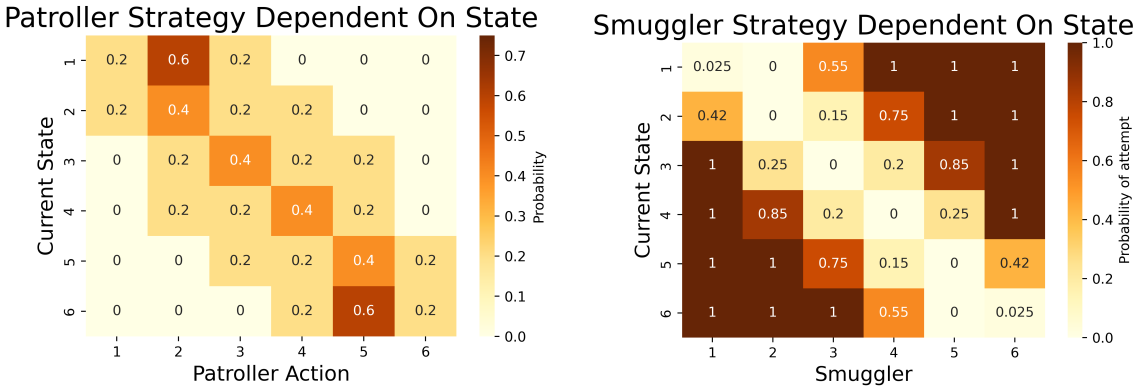
We consider a cost function of $C(a) = 4a$. The reward to each smuggler for success is just the amount of items they send, so that $r_i = 1$ for every location i. We define the movement cost for the patroller to be $m_{i,j} = |i - j|^2$. The number of locations in the border shall be varied to display how the methods scale with the size of the problem. Finally, we consider a fixed discount factor of $\gamma = 0.9$ for each player.

In Table 5.6.1, we present the time it takes for five different algorithms to find a Nash equilibrium in the model. The first method is to solve a single linear program using the formulation of Raghavan, 2003 for single-controller stochastic games. The other methods use the iterative method of Shapley, 1953 in Algorithm 5 with different methods to find the solution to the maximisation problem in Step 3. The first of these deploys a linear program using a formulation by Filar and Vrieze, 2012. Subsequent approaches solve it as a resource allocation problem using the algorithms of Fox, 1966 and Kaplan et al., 2019. The final method reported solves using our method assuming a linear cost function in Algorithm 6. We set the tolerance ϵ in Algorithm 4 to 10^{-3}, and the scaling of the resource allocation to $\delta = 0.2$. Note that since $K = n/\delta = 5n$, it is always divisible by $r_b + C(1) = 5$. Therefore, by Theorem 5.5.5 the resource allocation problem finds the optimal solution.

Table 5.6.1: Time taken (secs.) to solve Example 1 with different numbers of locations n

Algorithm Used	$n = 6$	$n = 9$	$n = 12$	$n = 15$
Single Controller Linear Program	2.50×10^{-1}	2.297	2.5563×10^{1}	$7.665\,78 \times 10^{2}$
Shapley method with linear programming	2.00	4.7125×10^{1}	$3.381\,88 \times 10^{2}$	$2.812\,219 \times 10^{3}$
Shapley with resource allocation (Fox 1966)	1.9016×10^{1}	5.2353×10^{1}	7.8516×10^{1}	$1.202\,81 \times 10^{2}$
Shapley with resource allocation (Kaplan et al. 2019)	1.9859×10^{1}	5.0000×10^{1}	$1.030\,94 \times 10^{2}$	$1.326\,09 \times 10^{2}$
Shapley with Algorithm 6	6.3×10^{-2}	1.25×10^{-1}	2.03×10^{-1}	2.81×10^{-1}

Table 5.6.1 shows that Algorithm 6 dramatically speeds up the calculation of a Nash equilibrium in our game having a 400%, 1800%, 13000% and 43000% improvement in each respective example over the next best method. We take the case with $n = 6$ locations and show the patroller's strategy for a Nash equilibrium in Figure 5.6.1.



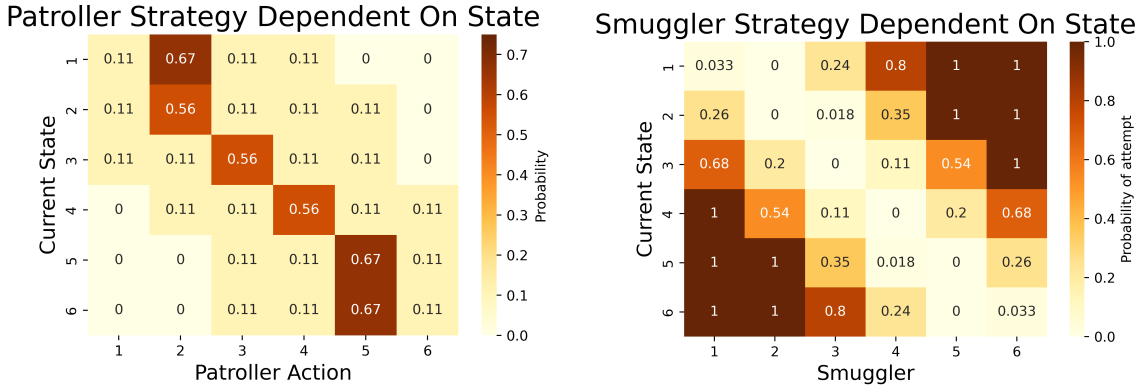
(a) (b)

Figure 5.6.1: A Nash equilibrium in Example 1. The vertical axis gives the current state s of the system in both figures. In (a) the horizontal axis shows each location the patroller could move to and the colour gives the probability with which they take that action. In (b) the horizontal axis gives each smuggler and the colour gives the probability with which they make an attempt to smuggle an item.

We see in Figure 5.6.1(a) an illustration of the result of Lemma 5.5.4 and Theorem 5.5.5, with the patroller choosing actions with probability in multiples of $0.2 = r_b/(r_b + C(1))$. Similarly in Figure 5.6.1(b) we see that the smuggler's best response is to send an item with probability one, with probability zero or an intermediate value if the location is protected with respectively a probability less than, greater than or exactly

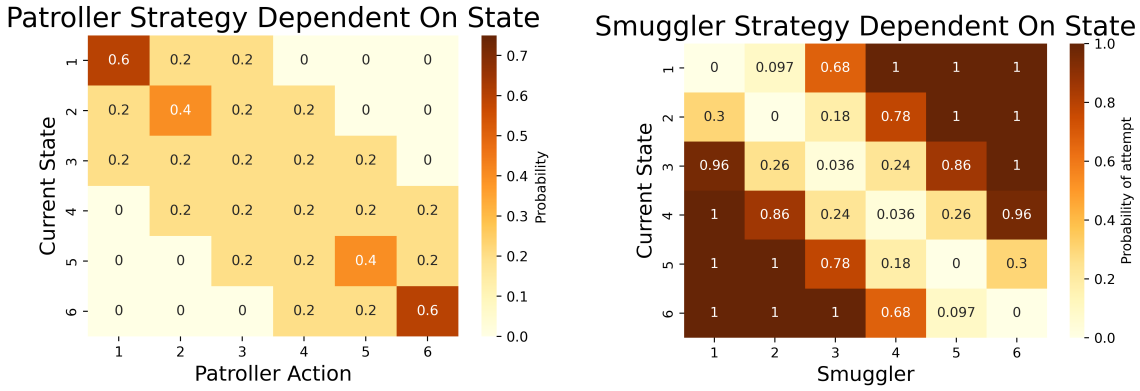
0.2. Note that as a consequence of Corollary 5.5.3, the results in Example 1 continue hold if we had a concave cost function C taking the values $C(0) = 0$ and $C(1) = 4$.

To illustrate how Nash equilibria in the game are dependent on the parameters of the model, we introduce two modified versions of Example 1, referred to as Examples 1.1 and 1.2. All parameters of the model in these examples are identical to those in Example 1 except that in Example 1.1 the cost function C is changed to $C(a) = 8a$, and in Example 1.2 the discount factor γ is reduced to 0.5. The Nash equilibria for Examples 1.1 and 1.2 can be found in Figures 5.6.2 and 5.6.3 respectively.



(a) (b)

Figure 5.6.2: A Nash equilibrium in Example 1.1. The vertical axis gives the current state s of the system in both figures. In (a) the horizontal axis shows each location the patroller could move to and the colour gives the probability with which they take that action. In (b) the horizontal axis gives each smuggler and the colour gives the probability with which they make an attempt to smuggle an item.



(a) (b)

Figure 5.6.3: A Nash equilibrium in Example 1.2. The vertical axis gives the current state s of the system in both figures. In (a) the horizontal axis shows each location the patroller could move to and the colour gives the probability with which they take that action. In (b) the horizontal axis gives each smuggler and the colour gives the probability with which they make an attempt to smuggle an item.

Figure 5.6.2 shows how increasing the values of $C(a)$ enables the patroller to prevent more items from being successfully smuggled. The patroller now only needs to defend a location with probability $r_b/(r_b + C(1)) = 1/9$ (as opposed to the value $1/5$ in Example 1) in order to prevent the smuggler from obtaining a strictly positive expected reward by sending an item.

By comparing Figure 5.6.3 with Figure 5.6.1 we may observe the effect of decreasing the discount factor on the Nash equilibrium. For example, consider the actions taken by the patroller when they are at location 1. In Example 1, the patroller moves to location 2 with probability 0.6. Indeed, state 2 has a higher value than state 1, which provides an incentive for the patroller to pay the movement cost. However, in Example 1.2 the patroller instead remains at location 1 with probability 0.6. Even though location 2 is still of higher value, the decrease in the discount factor implies that future rewards are less valuable.

5.6.2 Example 2: Linear Border With Strictly Convex Cost Function

We now give an example with a strictly convex cost function and show how this yields a different solution from the previous example. The parameters of the model are identical to those in Example 1 ($r_i = 1$ for all i, $m_{i,j} = |i-j|^2$, $\gamma = 0.9$), except now we take $C(a) = 4a^2$. Note that we still have $C(0) = 0$ and $C(1) = 4$ as before, and so in our results demonstrate that the simplifications afforded in Example 1 for the concave case no longer apply.

The strategies obtained in this subsection are not necessarily Nash equilibria, since by using the discretization of the resource allocation problem in (5.4.8) we derive only a δ-optimal solution. We assess the closeness to equilibrium by examining the worst case expected reward to the patroller under their strategy $\mathbf{\Pi}$. We calculate the worst case expected reward (WCER) by finding a strategy for the smugglers $\mathbf{\Xi}^*$ that is the best response to the patroller's strategy $\mathbf{\Pi}$. Assuming a uniform distribution over the inital state of the system, $\mathbb{P}(s_0 = s) = 1/n$, we can calculate the WCER for the patroller as follows.

$$\begin{aligned} WCER(\mathbf{\Pi}) &= \min_{\mathbf{\Xi} \in (\Delta([0,1]^n))^n} \left\{ \frac{1}{n} \sum_{s=1}^{n} \mathbb{E}_{\mathbf{\Pi},\mathbf{\Xi}} \left[\sum_{t=0}^{\infty} \gamma^t \tilde{R}_{pat}(b_t, \boldsymbol{a}_t, s_t) \,\middle|\, s_0 = s \right] \right\} \\ &= \frac{1}{n} \sum_{s=1}^{n} \mathbb{E}_{\mathbf{\Pi},\mathbf{\Xi}^*} \left[\sum_{t=0}^{\infty} \gamma^t \tilde{R}_{pat}(b_t, \boldsymbol{a}_t, s_t) \,\middle|\, s_0 = s \right]. \end{aligned}$$

Note that,

$$WCER(\mathbf{\Pi}^*) = \frac{1}{n} \sum_{s=1}^{n} \boldsymbol{V}_{pat}(s) \geq WCER(\mathbf{\Pi}).$$

The two methods that we implement to solve this example are the resource allocation algorithms of Fox, 1966 and Kaplan et al., 2019 within the iterative algorithm of Shapley, 1953. The resource allocation problem now becomes more challenging to solve, compared to the linear case, since the smugglers' best response to the patroller is more

complex. Therefore, there is no scaling of the continuous problem (5.4.7) that will give us the optimal solution. Now, the smaller the choice of δ, the better the strategy $\mathbf{\Pi}$ computed. In Table 5.6.2, we give the time taken and worst case reward for the two algorithms under different choices of scaling. As in Table 5.6.1, a tolerance of $\epsilon = 10^{-3}$ was used for Algorithm 5.

Table 5.6.2: Worst case expected reward in Example 2

	$\delta = 1$	$\delta = 0.2$	$\delta = 0.1$	$\delta = 0.04$
$n = 6$	-39.068	-38.338	-38.291	-38.282
$n = 9$	-67.740	-67.571	-67.551	-67.544
$n = 12$	-97.681	-97.239	-97.230	-97.227
$n = 15$	-127.200	-127.060	-127.052	-127.049

In Table 5.6.2, we can see that as δ gets smaller the worst case expected reward improves for the patroller. Tables 5.6.3 and 5.6.4 give the time taken to compute the strategies shown in Table 5.6.2.

Table 5.6.3: Time taken (s) to solve Example 2 (Fox 1966)

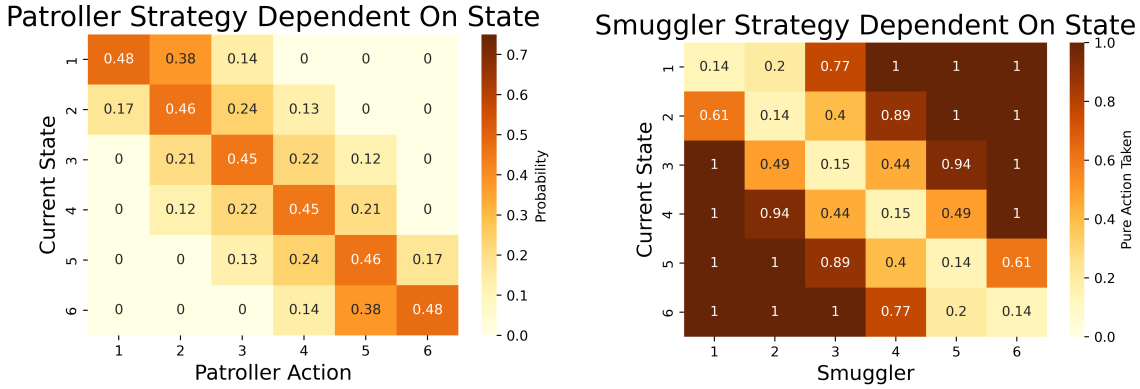
	$\delta = 1$	$\delta = 0.2$	$\delta = 0.1$	$\delta = 0.04$
$n = 6$	6.688	16.125	29.453	82.875
$n = 9$	17.484	44.844	84.281	192.391
$n = 12$	31.359	91.844	159.641	337.500
$n = 15$	45.375	139.984	603.641	1407.313

Table 5.6.4: Time taken (s) to solve Example 2 (Kaplan et al. 2019)

	$\delta = 1$	$\delta = 0.2$	$\delta = 0.1$	$\delta = 0.04$
$n = 6$	6.641	18.656	24.563	30.719
$n = 9$	19.672	50.469	69.016	83.156
$n = 12$	31.453	118.344	128.234	147.734
$n = 15$	49.844	143.313	498.938	591.016

In the cases with few locations and low fidelity of scaling, the algorithm by Fox, 1966 is quicker than that of Kaplan et al., 2019 but as the problem size grows this is

no longer the case. Note how in the six location example changing the scaling factor has a much bigger effect on the worst case expected reward of the strategy than in the fifteen location problem. In Figure 5.6.4, we show the strategy calculated for $\delta = 0.04$ and six locations.



(a) (b)

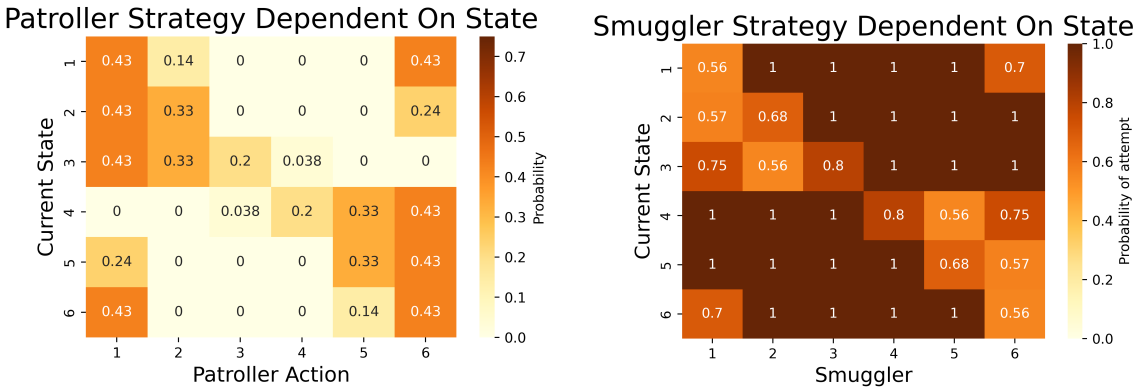
Figure 5.6.4: A Nash equilibrium in Example 2. The vertical axis gives the current state s of the system in both figures. Figure 5.6.4(a) has the same interpretation as in Figure 5.6.1(a). In (b) the horizontal axis denoted each smuggler and the colour now gives the quantity of items they attempt to smuggle with probability one.

We can see that the strategy given in Figure 5.6.4(a) is quite different to that in Figure 5.6.1. No longer are the probabilities multiples of $r_b/(r_b+C(1))$. The patroller is now less likely to move away from one of the two edges of the border, a key impact that changing the cost function has had on their decision making. In Figure 5.6.4(b), there is also a large difference in the strategy displayed compared to Figure 5.6.1(b), having a single action in $[0, 1]$ taken with probability one by each smuggler. These differences elucidate the importance we ascribe to the modelling of costs in our analysis.

5.6.3 Example 3: Perimeter Border With Linear Cost Function

We now turn our attention to an alternative border structure that is important operationally, namely a circular perimeter of an area. We now define the movement cost as

the minimum of the length of the two paths the patroller could take between locations. This yields $m_{i,j} = \min\{|i-j|, n-|i-j|\}$ for $i, j \in [n]$. We also consider a setup in which rewards for the smugglers are location dependent. In reality, there could be various reasons for this including the difficulty in getting through the border and the value of the items on the other side. Here, we set the rewards $\boldsymbol{r}$ equal to $(3, 2, 1, 1, 2, 3)$. The remaining parameters of the model those of the first example ($C(a) = 4a$, $\gamma = 0.9$). Equilibria in this example are computed as for Example 1.



(a) (b)

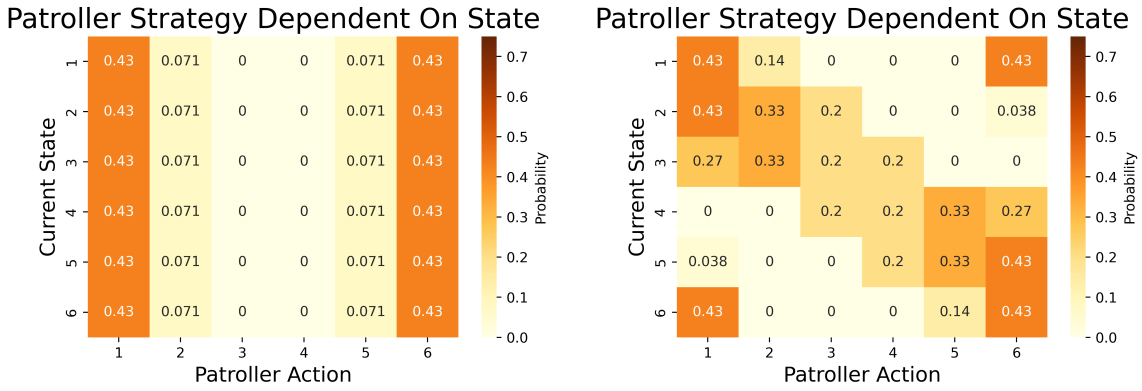
Figure 5.6.5: A Nash equilibrium in Example 3. The figure has the same interpretation as Figure 5.6.1.

Figure 5.6.5(a) shows how the locations protected most heavily are those with higher smuggler reward, which in this example are locations one and six. Note that since the locations form a circle, the patroller can travel from location one to location six at a cost of one unit. This is one reason why Figure 5.6.5(a) looks different from the patroller strategies in previous examples. We continue with the pattern of Figure 5.6.1(a), namely that the patroller protects location b with probability $r_b/(r_b + C(1))$. The other values in Figure 5.6.5(a) arise as a result of the probabilities needing to sum to one.

5.6.4 Value Of Modelling as a Stochastic Game

In this subsection, we will evaluate the benefits of using our model over alternative modelling approaches. We consider how the patroller's worst case expected reward would be affected if the game was considered to be normal-form, with no consideration to the state of the system in the next time step. We achieve this by finding the patroller's strategy in a Nash equilibrium when we set the discount rate to γ to zero.

The first case we consider is one in which all movement costs are set to zero. This means that the state of the system no longer has an effect on the rewards to either player. We make this choice since the assumptions in the work of Pita et al., 2008 or Alpern et al., 2011 are similar. However, the patroller will be accumulating costs to travel without knowing of their existence. To make a fairer comparison, we consider a second case where movement costs are considered by the patroller but she still acts myopically. This is equivalent to solving a normal-form game for each state in which the patroller could start.



(a) (b)

Figure 5.6.6: Patroller's strategies in a Nash equilibrium for the two described models. The figures gave the same interpretation as Figure 5.6.1(a).

In Figure 5.6.6, we show the strategies obtained under these two sets of assumptions for the model given in Example 3. Figure 5.6.6(a) shows why it is important to consider the geography underlying the model, with the patroller making large moves at a high

cost. In Figure 5.6.6(b) we overcome this, but the patroller can still be seen making suboptimal moves because she is not accounting for the value of the state to which she moves. In Table 5.6.5, we take the six location version of the three examples introduced previously in the section, find the strategies as detailed in the previous paragraph and calculate the worst case expected reward for the patroller in each case.

Table 5.6.5: Worst Case Expected Rewards Under A Range of Models

Model	Example 1	Example 2	Example 3
Normal-form Game (without movement cost)	-68.333	-73.958	-64.238
Normal-form Game (with movement cost)	-34.000	-38.743	-61.189
Stochastic Game	-33.587	-38.282	-60.110

Having no consideration for the states in the game leads to a large decrease in the reward to the patroller. In Example 1, for example, she incurs over twice the cost than in the full model. Factoring in movement costs but still disregarding future rewards improves the outcome to the patroller, but there is still a very significant benefit to solving with the full stochastic game model.The computational challenge of developing solutions may have been grounds for the earlier focus on over-simple models. Our analysis removes many of these obstacles.

5.7 Conclusion

In this chapter, we have introduced a new model which can be used to consider the interaction between smugglers and a patroller on a border. A number of properties of Nash equilibria in the stochastic game are established, and new algorithms to find these equilibria are developed. We provide examples to show empirically that our methods solve the model quicker than existing methods and additionally that using a stochastic game formulation achieves significant improvement for the patroller.

There are a number of ways that the model in this chapter could be extended,

including the addition of more patrollers, the ability for patrollers to catch more than a single smuggler and further constraints on quantities of items that may be smuggled. These extensions represent possible avenues for future work.

Chapter 6

Multiple Patroller Extension

This chapter extends the results of Chapter 5 by incorporating multiple patrollers to the model. The solution methods in Chapter 5 are not applicable to this new model, and therefore new approaches must be considered.

6.1 Introduction

The stochastic game framework for patrolling a border introduced in the previous chapter is restrictive in that the analysis only holds for one patroller. In this chapter we extend the framework to multiple patrollers, explain why new methodology is required, and consider three different approaches to the multiple patroller problem.

Firstly, we look to find methods which can exactly find the Nash equilibria in the multiple patroller game. We use a subgradient descent method to find the strategies in a Nash equilibria, and consequently, the state values of the game. We look at how such an approach can be implemented, and then go on to show how it performs.

Secondly, we introduce two heuristics which we think are good approximations to the Nash equilibria. Our first heuristic assumes that the patrollers are myopic and do not value future rewards. Myopic policies are easy to compute, since we do not need to take into consideration how the immediate actions will affect the value of future turns.

Our second heuristic partitions the border into separate sections, each section being defended by one patroller. The strategy used by a patroller on their section can then be computed by the analysis in the previous section, which we have already shown to be computationally efficient.

Finally, we apply existing reinforcement learning approaches to our model. The reinforcement learning algorithms we try have been introduced in Chapter 2 of the thesis, and consist of a fictitious play and a Q-learning approach. We provide a numerical study of their performance on our problem, and discuss some challenges with implementing them.

6.2 Model Description

In this section we extend the model of the previous chapter to allow for multiple patrollers. We introduce the state of the game, the action spaces of the players and define a Nash equilibrium in our model.

We consider a border made up of n locations labelled from 1 to n inclusive. We will use the notation $[n]$ to denote the set of all locations where $[n] = \{1, \dots, n\}$. Time will be modelled in discrete steps $t = 0, 1, \dots$. Such time steps are natural here, where decisions could be taken on an hourly or daily basis.

We present the model in this section by taking the smugglers collectively to be a single player. The equilibria in the game where smugglers are aggregated as a single player or are considered individually coincide. The result is identical to Lemma 5.3.2, and we omit the proof as it follows from a similar statement found in Filar (1985). Additionally, we also present the multiple patrollers as a single player, which we can think of as having some central controller who instructs them where to move on the border.

Thus, we look to define a stochastic game between two players: the k patrollers and

the n smugglers. The patrollers begin each time step t at some locations $\boldsymbol{s} = (s_t^1, \ldots, s_t^k)$, which we take to be the current state of the system. Hence, the state space of the game is $\mathcal{S} = [n]^k$. The patrollers pick a vector of locations $\boldsymbol{b}_t = (b_t^1, \ldots, b_t^k)$ to defend, and the smugglers pick a quantity of items from the interval $[0, 1]$ to send to each location. We write the smugglers' action as $\boldsymbol{a}_t = (a_t^1, \ldots a_t^n)$, where a_t^i is the quantity sent to the location i. Note that the assumption of actions in the unit interval is without loss of generality, since we can account for quantities from the interval $[0, q]$ for some $q > 0$ by a scaling of the actions. Hence, the action space of the patrollers and smugglers respectively at each epoch are $\mathcal{A}_{pat} = [n]^k$ and $\mathcal{A}_{smug} = [0, 1]^n$. Both the patrollers and the smugglers take an action simultaneously, with no knowledge of the action chosen by the opponent. The state of the system at the next time step is the previous action of the patroller, and so

$$\mathbb{P}(\boldsymbol{s}_{t+1} = \boldsymbol{b} \mid \boldsymbol{b}_t = \boldsymbol{b}, \boldsymbol{a}_t = \boldsymbol{a}, \boldsymbol{s}_t = \boldsymbol{s}) = \mathbb{P}(\boldsymbol{s}_{t+1} = \boldsymbol{b} \mid \boldsymbol{b}_t = \boldsymbol{b}) = 1 \tag{6.2.1}$$

for all $\boldsymbol{b} \in \mathcal{A}_{pat}$, $\boldsymbol{a} \in \mathcal{A}_{smug}$ and $\boldsymbol{s} \in \mathcal{S}$. As we will see, the players can choose their actions according to some probability distribution which results in a random state transition in the game.

The patrollers catch all items sent by the smugglers to the locations they have chosen to defend. Therefore, there is no benefit for the patrollers having more than one patroller at a single location. At every other location, the items are successfully smuggled. We define $\mathcal{U}(\boldsymbol{b})$ to be the unique elements of the vector $\boldsymbol{b}$, corresponding to the set of locations which are under protection. Smugglers receive a fixed reward of $r_i > 0$ for each unit of item smuggled through the location i. However, if caught, the smugglers must pay a penalty related to the amount smuggled. This is determined by the cost function $C : [0, 1] \to \mathbb{R}_+$. We assume that C is an increasing function with $C(0) = 0$. The patrollers' payoff is equal to the negative of the smugglers' payoff, but they must additionally pay a cost for each patroller moving from one location to

another. These movement costs are given by the parameters $m_{i,j} \geq 0$ $i, j \in [n]$. Thus, the reward functions of the patrollers and the smugglers respectively are as follows:

$$R_{pat}(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{s}) = \sum_{i \in \mathcal{U}(\boldsymbol{b})} C(a_i) - \sum_{i \in [n] \setminus \mathcal{U}(\boldsymbol{b})} r_i a_i - \sum_{j=1}^{k} m^j_{\boldsymbol{s},\boldsymbol{b}}$$

$$R_{smug}(\boldsymbol{b}, \boldsymbol{a}) = \sum_{i \in [n] \setminus \mathcal{U}(\boldsymbol{b})} r_i a_i - \sum_{i \in \mathcal{U}(\boldsymbol{b})} C(a_i).$$

The game continues for an infinite number of time steps, with rewards discounted at a rate of $\gamma \in [0, 1)$ for the players.

A pure action is an action which a player is able to perform. In our case these are the elements of the sets $\mathcal{A}_{pat}$ and $\mathcal{A}_{smug}$ for the patrollers and smugglers respectively. Instead of picking a pure action deterministically, players can draw an action at random, according to a probability distribution over their pure actions. A stationary mixed strategy for either player is an n-tuple of probability distributions over the pure actions of a player,

$$\boldsymbol{\Pi} = (\boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^n) \in (\Delta([n]^k))^n$$

$$\boldsymbol{\Xi} = (\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^n) \in (\Delta([0, 1])^n)^n.$$

The strategies for the patrollers and smugglers respectively given that the state of the system is i are $\boldsymbol{\pi}^i$ and $\boldsymbol{\xi}^i$. Assuming the strategies are fixed over time, we write the expected discounted reward for both players over an infinite horizon as

$$U_{pat}(\boldsymbol{\Pi}, \boldsymbol{\Xi}) = \mathbb{E}_{\boldsymbol{\Pi}, \boldsymbol{\Xi}, \mathbb{P}_0} \left[\sum_{t=0}^{\infty} \gamma^t R_{pat}(\boldsymbol{b}_t, \boldsymbol{a}_t, \boldsymbol{s}_t) \right],$$

and,

$$U_{smug}(\boldsymbol{\Pi}, \boldsymbol{\Xi}) = \mathbb{E}_{\boldsymbol{\Pi}, \boldsymbol{\Xi}, \mathbb{P}_0} \left[\sum_{t=0}^{\infty} \gamma^t R_{smug}(\boldsymbol{b}_t, \boldsymbol{a}_t) \right]. \tag{6.2.2}$$

In (6.2.2) expectations are taken with respect to the strategies of both players so that

$\boldsymbol{b}_t \sim \boldsymbol{\pi}^{\boldsymbol{s}_t}$ and $\boldsymbol{a}_t \sim \boldsymbol{\xi}^{\boldsymbol{s}_t}$, and also with respect to the probability distribution $\mathbb{P}_0$ over the initial state $\boldsymbol{s}_0$. Since the outcome of one player depends on the action of the other, it is not possible to maximize the rewards of the players independently. We give the definition of a Nash equilibrium as first given by Nash (1950).

Definition 6.2.1. *The strategies $\boldsymbol{\Pi}^*$ and $\boldsymbol{\Xi}^*$ for the patrollers and smugglers respectively form a Nash equilibrium for the game if and only if,*

$$U_{pat}(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*) \geq U_{pat}(\boldsymbol{\Pi}, \boldsymbol{\Xi}^*) \; \forall \; \boldsymbol{\Pi} \in (\Delta([n]^k))^n$$
$$U_{smug}(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*) \geq U_{smug}(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}) \; \forall \; \boldsymbol{\Xi} \in (\Delta([0,1]^n))^n.$$

Nash equilibria give the most natural solution for our model, in that they provide the best possible lower bound of the discounted expected reward to the patrollers. This could be operationally important if, for example, the smugglers were to discover the strategy of the patrollers and were able to optimize their strategy using this knowledge.

6.3 Finding Nash Equilibria

In the previous chapter, an algorithm is developed which can compute the Nash equilibria efficiently in the one patroller version of the framework. In this section, we discuss why the analysis made in the chapter does not hold in the framework with multiple patrollers.

There exist in the literature algorithms that can calculate Nash equilibria in two-player zero-sum stochastic games, such as the one by Shapley (1953). However, their assumption that the game is finite means that they are not directly applicable here. In this section, we present a method for determining Nash equilibria in our game. We begin by defining the value of a state s for the players.

Definition 6.3.1. *The value $\boldsymbol{V}_{pat}(\boldsymbol{s})$ of a state $\boldsymbol{s}$ to the patrollers in the stochastic*

game is the expected reward to the patrollers in a Nash equilibrium $(\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*)$*, given that the system starts in the state* $\boldsymbol{s}$*, namely*

$$\boldsymbol{V}_{pat}(\boldsymbol{s}) = \mathbb{E}_{\boldsymbol{\Pi}^*, \boldsymbol{\Xi}^*}\left[\sum_{t=0}^{\infty} \gamma^t R_{pat}(\boldsymbol{b}_t, \boldsymbol{a}_t, \boldsymbol{s}_t) \,\middle|\, \boldsymbol{s}_0 = \boldsymbol{s}\right].$$

The value of a state $\boldsymbol{s}$ *for the smugglers,* $\boldsymbol{V}_{smug}(\boldsymbol{s})$*, is defined similarly.*

The value of each state is unique and can be seen to solve the system of equations,

$$\boldsymbol{V}_{pat}(\boldsymbol{s}) = \max_{\boldsymbol{\pi} \in \Delta([n]^k)} \min_{\boldsymbol{a} \in [0,1]^n} \left[\sum_{\boldsymbol{b} \in [n]^k} \pi_{\boldsymbol{b}} \left\{R_{pat}(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{s}) + \gamma \boldsymbol{V}_{pat}(\boldsymbol{b})\right\}\right] \tag{6.3.3}$$

where $\pi_{\boldsymbol{b}}$ is the probability the patrollers take a joint action $\boldsymbol{b}$. This follows from Shapley (1953) and Maitra and Parthasarathy (1970). By (6.2.1), the transitions of system state are determined entirely by the patrollers' choice of action. This is why in (6.3.3) we can deterministically know the system state resulting from any action taken by the patrollers.

Shapley (1953) proved that given any initial starting values $\{\boldsymbol{V}^0_{pat}(\boldsymbol{s}) \mid \boldsymbol{s} \in \mathcal{S}\}$ the sequence $\{\boldsymbol{V}^k_{pat}(\boldsymbol{s}) \mid \boldsymbol{s} \in \mathcal{S}\}_{k=1}^{\infty}$, determined by the recursion

$$\boldsymbol{V}^k_{pat}(\boldsymbol{s}) = \max_{\boldsymbol{\pi} \in \Delta([n]^k)} \min_{\boldsymbol{a} \in [0,1]^n} \left[\sum_{\boldsymbol{b} \in [n]^k} \pi_{\boldsymbol{b}} \left\{R_{pat}(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{s}) + \gamma \boldsymbol{V}^{k-1}_{pat}(\boldsymbol{b})\right\}\right]$$

converges to $\{\boldsymbol{V}_{pat}(\boldsymbol{s}) \mid \boldsymbol{s} \in \mathcal{S}\}$ as $k \to \infty$. When state and action spaces are finite, state values may be obtained by using linear programming to solve the maximization problem within Shapley's iteration. However, since we assume that the action space of the smugglers is infinite, this approach is not available to us. Therefore, we look elsewhere to solve (6.3.3).

We begin by establishing properties of the smugglers' best response against any strategy taken by the patrollers. If the smugglers take a best response against the

patrollers' strategy $\boldsymbol{\pi}$ when the system state is $\boldsymbol{s}$, the patrollers receive a payoff which we shall denote as $G(\boldsymbol{\pi}, \boldsymbol{s}, \boldsymbol{V}_{pat})$, where

$$G(\boldsymbol{\pi}, \boldsymbol{s}, \boldsymbol{V}_{pat}) = \min_{\boldsymbol{a} \in [0,1]^n} \left\{ \sum_{\boldsymbol{b} \in [n]^k} \pi_{\boldsymbol{b}} [R_{pat}(\boldsymbol{b}, \boldsymbol{a}, \boldsymbol{s}) + \gamma \boldsymbol{V}_{pat}(\boldsymbol{b})] \right\}.$$

If $\boldsymbol{V}_{pat}$ is the value function for the patroller, it will solve the following system of equations by (6.3.3).

$$\boldsymbol{V}_{pat}(\boldsymbol{s}) = \max_{\boldsymbol{\pi} \in \Delta([n])} G(\boldsymbol{\pi}, \boldsymbol{s}, \boldsymbol{V}_{pat}) \text{ for all } \boldsymbol{s} \in [n]^k.$$

In the previous chapter, the analysis then continues from the observation that with one patroller the function G is additively separable with respect to $\boldsymbol{\pi}$, and so we can write

$$G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}) = \sum_{b \in [n]} g_b(\pi_b, s, \boldsymbol{V}_{pat})$$

where

$$g_b(\pi_b, s, \boldsymbol{V}_{pat}) = - \max_{a \in [0,1]} \{(1 - \pi_b) r_b a - \pi_b C(a)\} + \pi_b(\gamma \boldsymbol{V}_{pat}(b) - m_{s,b})$$

However, with multiple patrollers we can no longer make this separation. As an example, consider a smuggler at some location i. In the one patroller problem, smuggler i's best response only depends on the probability location i is defended given by π_i and therefore we can separate the function G. In the multiple patroller problem, smuggler i's best response again depends just on the probability that location i is defended against. However, the probability of location i being defended against now depends on the probability of any joint action which includes patrolling location i. For example, in a two location problem we have that the probability of location i being defended is equal to,

$$[\pi_(i,1) + \ldots, +\pi_(i,n)] + [\pi_(1,i) + \ldots, +\pi_(n,i)] - \pi_(i,i).$$

Therefore, we are unable to separate the function G into the actions of the patroller.

6.4 Subgradient Descent

In the previous section, we showed that we could not maximize the function G as in the previous chapter. There are algorithms in the literature which we could apply to the stochastic game to find Nash equilibria. For example, Raghavan (2003) introduces a linear program which provides the solution to any two player single controller zero-sum stochastic game. Alternatively, we could use a linear program to find the value of every state at each iteration of the value iteration algorithm using the work of Shapley (1953) and Vrieze and Tijs (1982). However, both these methods have a common problem in that they do not scale computationally efficiently with the size of the problem. Since the algorithms require solving a maxi-min optimization problem which considers every possible action of both sides, it quickly becomes infeasible from a computational perspective to construct the linear programs with either algorithm. For example, the number of variables quickly increases as the total number of actions which could be played is equal to $n^k 2^n$. To overcome this problem, we introduce a subgradient descent method to replace the linear program. We show numerically that using the method finds a close approximation to a Nash equilibrium.

A subgradient is defined as follows.

Definition 6.4.1 (Subgradient)**.** *A subgradient of a convex function $f : I \to \mathbb{R}$, where I is an open interval, at a point $x_0 \in I$ is a number $c \in \mathbb{R}$ such that,*

$$f(x) - f(x_0) \geq c(x - x_0) \; \forall \; x_0 \in I$$

Note that if the function f is differentiable, we have that the subgradient and derivative coincide. We can find subgradients of functions which are not differentiable. For example, a piecewise linear function is not differentiable at one point, but there exist a set of subgradients at that point.

We implement the adaptive Polyak step size method of gradient descent found in Hazan and Kakade (2019) found in Algorithm 7. The algorithm has a set number K of epochs, where each epoch has T steps. At each step, a step size α_t is calculated depending on how far the algorithm estimates the current best solution is to the optimum. The algorithm makes a step in the direction determined by the subgradients with step size α_t. The subgradients can be easily calculated since the function G is piecewise linear. We need to constrain the solution to be a probability distribution since we are trying to find a strategy which maximizes the function G. Therefore, we project the solution from the subgradient descent method onto the probability simplex with Algorithm 10 from Wang and Carreira-Perpinán (2013), which can be found in the Appendix of this chapter. We denote the projection onto the probability simplex as $\mathcal{P}$. After each epoch, the estimated optimum of the maximization of the problem is updated to be the average of the current best solution and current estimated optimum.

Algorithm 7: Adaptive Polyak Algorithm (Hazan and Kakade, 2019)

Input: time horizon T, number of epochs K, starting value π_0, estimate

$$\tilde{G}_0 \geq f(G^*)$$

for *epoch* $k = 0, \ldots, K-1$ **do**

  for $t = 0, \ldots, T-1$ **do**

$$\alpha_t = \frac{G(\boldsymbol{\pi}_t) - \tilde{G}_k}{2\|\nabla G(\pi_t)\|_2^2}$$

$$\boldsymbol{\pi}_{t+1} = \mathcal{P}(\pi_t - \alpha_t \nabla G(\pi_t))$$

$$\tilde{G}_{k+1} = \frac{\tilde{G}_k + G(\pi_{T-1})}{2}$$

Output: π_{T-1}

The adaptive Polyak algorithm allows us to maximize the function G efficiently. The time horizon T and number of epochs K are hyperparameters of the algorithm which are decided before it runs. Recall that maximizing the function G gives us the value of the normal-form game given a fixed system state $\boldsymbol{s}$ and state values $\boldsymbol{V}_{pat}$. We can therefore now use the value iteration algorithm of Shapley (1953) in order to iterate through the state values as found in Algorithm 8, and find the Nash equilibrium of the stochastic game.

Algorithm 8: Calculation of state values

Input: $\epsilon > 0$

Initialise: $\boldsymbol{V}_{pat}^{0}(s) = (0, \ldots, 0)$ and $k = 1$

while $\max_{s \in \mathcal{S}} \left\{ \left| \boldsymbol{V}_{pat}^{k-1}(s) - \boldsymbol{V}_{pat}^{k}(s) \right| \right\} > \epsilon$ **do**

  for $s = 1, \ldots, n$ **do**

    Find,

$$\boldsymbol{V}_{pat}^{k}(s) := \max_{\boldsymbol{\pi} \in \Delta([n])} \min_{\boldsymbol{a} \in [0,1]^n} \left[\sum_{b=1}^{n} \pi_b \left\{ \tilde{R}_{pat}(b, \boldsymbol{a}, s) + \gamma \boldsymbol{V}_{pat}^{k-1}(b) \right\} \right]$$

$$= \max_{\boldsymbol{\pi} \in \Delta([n])} G(\boldsymbol{\pi}, s, \boldsymbol{V}_{pat}^{k-1})$$

    using the adaptive Polyak algorithm.

  end

  $k := k + 1$

end

Output: $\boldsymbol{V}_{pat}^{k}$

Note that the value iteration algorithm stores the value of each state, and additionally, we can remember the policy $\boldsymbol{\pi}^k$ which achieves this maximum. Therefore, with each iteration of Algorithm 8 we know better starting values for the subgradient method in Algorithm 7.

6.4.1 Numerical Example

In Figure 6.4.1 we observe the time of each iteration for a two patroller model, with parameters $r_i = 1$ and $C_i = 4$ for all $i \in [n]$, $m_{i,j} = (i - j)^2$ and $\gamma = 0.75$, on borders with $n = 10, 15$ and 20 locations.

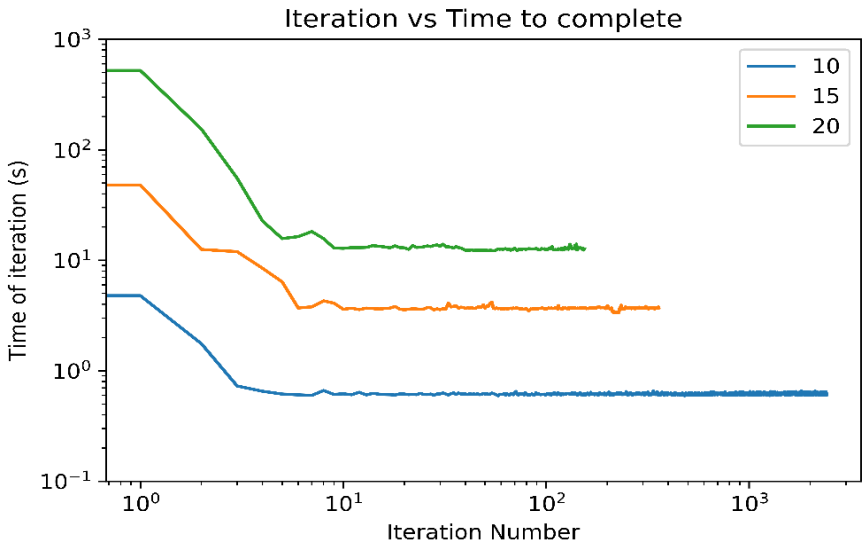


Figure 6.4.1: Time taken for iterations of the value iteration algorithm to be completed

We can see in Figure 6.4.1 that the iterations initially speed up, before eventually converging to a constant amount of time. Since the starting point of the subgradient method is given by the output of the last iteration, we conjecture that this is why the iterations get progressively faster. However, the speed cannot forever increase and so there reaches a constant amount of time needed to compute all the steps in the subgradient method, since the number of epochs and time horizon is fixed. A future direction for research on the problem could be to consider an algorithm which varies the length of each epoch and the time horizon depending on the current accuracy of the solution.

Given policy $\boldsymbol{\pi}$, we can evaluate its performance by taking the smugglers' best response $\boldsymbol{\xi}$ and calculating the expected reward to the patroller. We denote the performance for a policy $\boldsymbol{\pi}$ by $\tilde{V}^{\boldsymbol{\pi}}$. To study the convergence of the algorithm, we consider the mean difference between the performance of policies in successive iterations given by,

$$\frac{1}{|\mathcal{S}|}\sum_{\boldsymbol{s}\in\mathcal{S}}|V^{\boldsymbol{\pi}^k}(\boldsymbol{s}) - V^{\boldsymbol{\pi}^{k+1}}(\boldsymbol{s})| \tag{6.4.4}$$

We use the metric of performance found in (6.4.4) to evaluate the convergence of Algorithm 8 in Figure 6.4.2. We use the same model parameters as earlier in this section. The problem instances with 10 and 15 locations were given 1500 seconds to run, whilst the 20 location instance was given 3000.

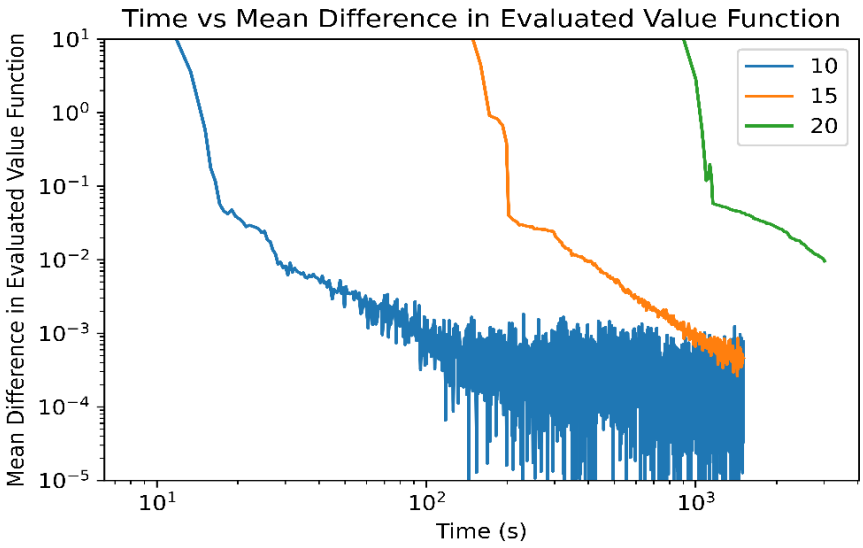


Figure 6.4.2: Time taken against the difference in evaluated value function for problem instances with differing numbers of locations

We can see in Figure 6.4.2 how on a logarithmic scale of time the convergence appears to follow a similar pattern. In each of the problem instances, there is a period of quick convergence with the initial slower iterations, before we reach faster iterations with smaller improvements on the policy's performance. Note that due to the logarithmic scale on the y-axis, the oscillations seen in 6.4.2 are very small.

In this section, we have introduced an algorithm based on subgradient descent to find Nash equilibria in the multiple patroller problem. However, as it is infeasible to calculate exact solutions in these games, we need a way to evaluate the performance of the algorithms. Therefore, we look to create heuristics which we think will give us a close approximation to the Nash equilibria in our model.

6.5 Heuristics

In this section, we discuss two intuitive heuristics which can provide strategies for the patrollers in the multiple patroller border patrol game. We can use the heuristics to evaluate the performance of the subgradient method when exact methods are too computationally expensive. We conclude the section by providing a numerical analysis of the heuristics' behavior on the example that we discussed in the previous section.

The first heuristic which we will discuss is a myopic policy for the patrollers. If the patrollers are acting myopically, they do not take into consideration the future value of any rewards they will receive. We can ignore the value of future states and instead just maximize their immediate reward function, simplifying computation. The heuristic lacks the ability to leave the patrollers in favorable positions for the next time step, for example, by placing them closer together than would be optimal.

The second heuristic considered is to partition the border into smaller sections, which are each protected only by one fixed patroller. We can then treat each border section as a one patroller problem, which we solved in the previous chapter. We will mainly focus on line and circle borders where we conjecture splitting the border into equal lengths is optimal, however, for more complex borders the decision of how to best partition the border would be more complex. By introducing 'fake edges' into the border which the patrollers cannot cross, the heuristic loses performance compared to the optimal solution.

6.5.1 Myopic Patroller

The myopic policy for the patrollers can be found by solving the stochastic game, but taking the discount factor γ to be zero. However, if we use other methods such as those previously discussed (Shapley (1953), Raghavan (2003)) then the same computation problems occur from there being $n^k 2^n$ possible actions, which quickly becomes infeasible

to solve. Instead, we conjecture an algorithm to find myopic policies for the patrollers, which follows a similar method to the solution method found in the previous chapter.

Suppose that the patrollers have a strategy $\tilde{\boldsymbol{\pi}}$, allowing the strategy to have probabilities which sum to less than one. For any action $\boldsymbol{b}$ which the patrollers could take, we can calculate the incremental change in their expected reward if they increased the probability of taking the action by some small $\delta > 0$. We define this incremental change in expected reward to be $g_{\boldsymbol{b}}(\boldsymbol{s}, \delta)$. Starting from a strategy of $\tilde{\boldsymbol{\pi}} = \mathbf{0}$, we propose that the patrollers greedily increase the probability of taking an action in the set,

$$\arg\max_{\boldsymbol{b} \in [n]^k} \left\{ g_{\boldsymbol{b}}(\boldsymbol{s}, \delta) \right\}.$$

We define the algorithm in full below.

Algorithm 9: Find myopic policy

Initialise: $\tilde{\boldsymbol{\pi}} = \mathbf{0}, k = 0$

while $\sum \tilde{\boldsymbol{\pi}} < 1$ **do**

  Let,

$$j \in \arg\max_{b \in [n]} \left\{ g_b \left(\tilde{\boldsymbol{\pi}}, k \right) \right\}$$

  with ties decided by taking the lowest index.

  $\tilde{\pi}_b := \tilde{\pi}_b + k$

end

Output: $\tilde{\boldsymbol{\pi}}_K$

Although we have no analytical proof the output of the algorithms is the optimal myopic policy, when compared against a number of examples which can be exactly solved it gave the correct solution. Proving that Algorithm 9 always finds the myopic policy is an open question for future research.

6.5.2 Partitioning the Border

The second heuristic for the multiple patroller border patrol problem is to partition the border into multiple segments with one patroller each.

It is an open question of how best to partition the border. Intuitively, on a line border if the parameters of the game are equal across the locations, then we would split the border into equal lengths. However, if the border was a more complex shape or had locations that are more valuable for the smugglers to attack, it becomes less clear how the partition should be created. Another possibility would be to have the borders split into segments which overlap, however, we have not yet looked into this possibility.

Once there has been a partition established, we can then use the methods found in the previous chapter to find the patroller's strategy in each segment. Then we can take the joint distribution across each strategy to find the joint strategy of all patrollers. Note that since the segments are partitioned into non-overlapping sections, it does not matter how we take the joint distribution.

6.5.3 Numerical Example

We now compare the convergence of the subgradient method against the two heuristics introduced in this section.

The parameters of the game we will consider are $n = 10$ and $n = 20$ locations, $k = 2$ patrollers, $r_i = 1$ and $C_i = 4$ for all $i \in [n]$, $m_{i,j} = (i - j)^2$ and $\gamma = 0.75$. The problem instance with $n = 10$ locations will be given 60 seconds to solve, whilst the problem instance with 20 locations will be given 7200 seconds.

The ten and twenty location problems were given 60 and 7200 seconds respectively to run.

Our metric for the performance of an algorithm will be to take the patrollers' policy, calculate the smugglers best response to this, and then find the expected reward to them given some initial starting state. We will look at the expected reward from two differing

initial system states, either the patrollers begin at the edges of the border or from the middle of the border.

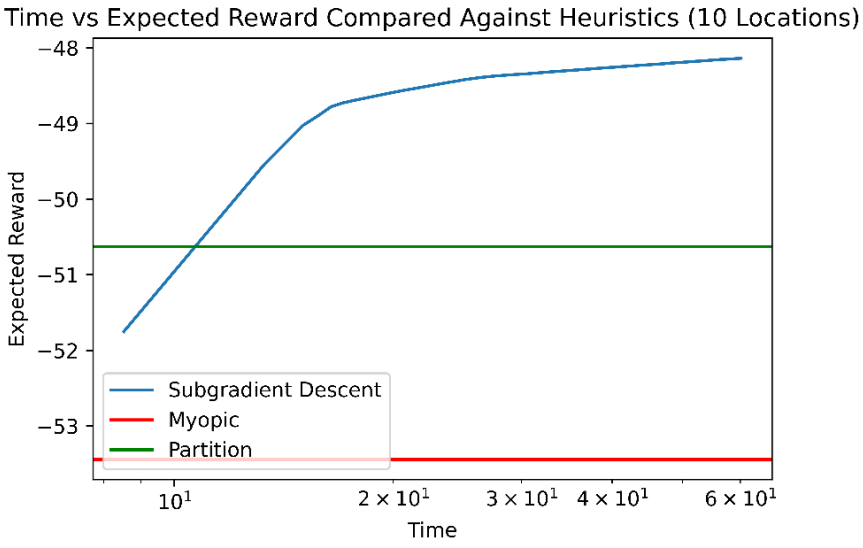


Figure 6.5.3: Convergence of policy performance with ten locations when patrollers begin at the middle of the border

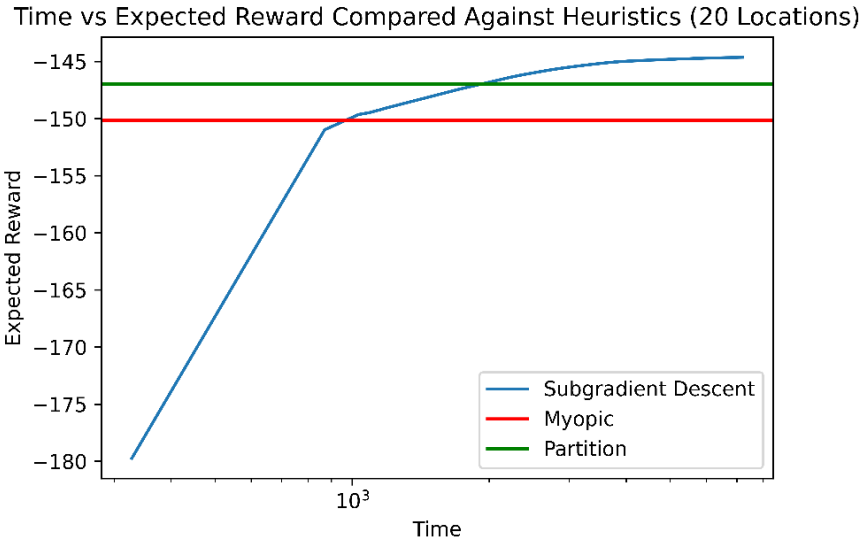


Figure 6.5.4: Convergence of policy performance with twenty locations when patrollers begin at the middle of the border

We can see in Figure 6.5.3 that with ten locations, the subgradient method quickly outperforms the two heuristics we have implemented. However, in Figure 6.5.4 we can

see on the twenty location problem, the first iteration of the subgradient method takes a long time to complete. Note, we are initializing the subgradient method from an arbitrary starting point with state values equal to zero and uniform strategies across the patrollers' actions.

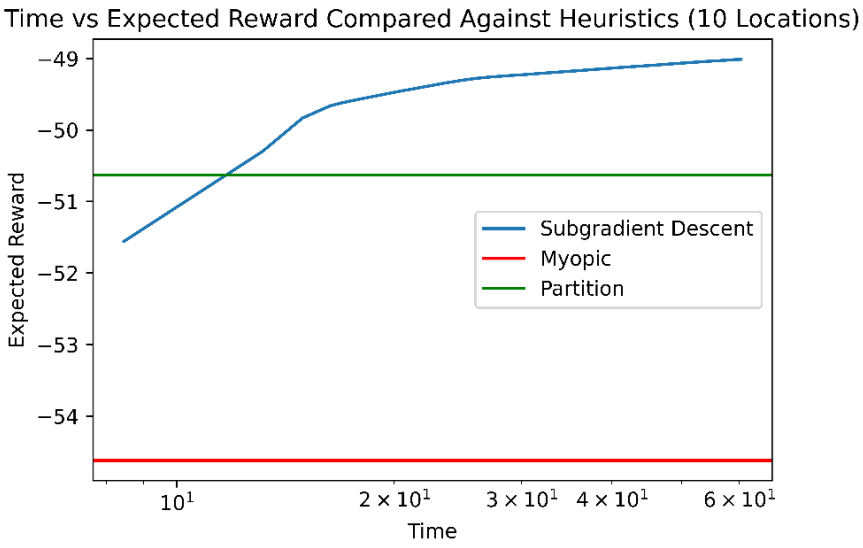


Figure 6.5.5: Convergence of policy performance with ten locations when patrollers begin at opposite ends of the border

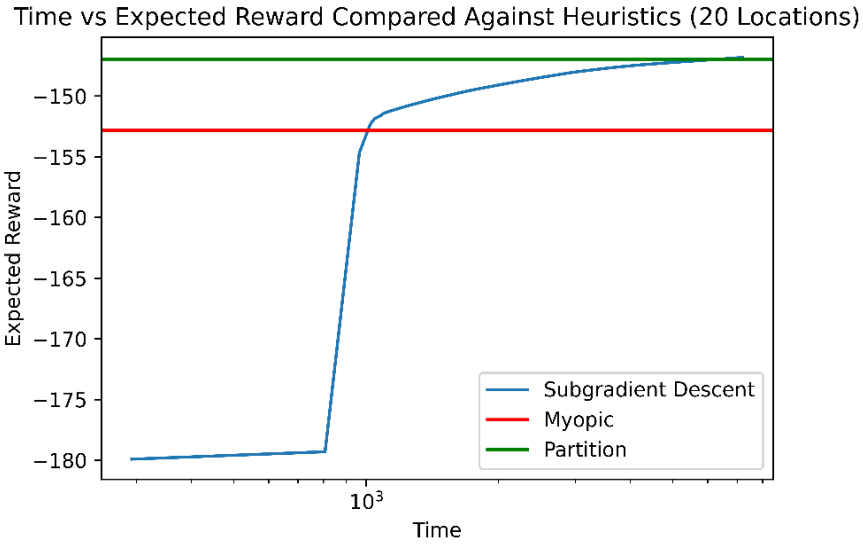


Figure 6.5.6: Convergence of policy performance with twenty locations when patrollers begin at opposite ends of the border

In Figures 6.5.5 and 6.5.6 we see the partition heuristic is performing better than in Figures 6.5.3 and 6.5.4 when starting in the middle. The behavior is expected since when starting at the edges of the border, the impact of the patrollers potentially meeting is small with a small discount factor ($\gamma = 0.75$). Again, the subgradient method quickly outperforms the heuristics in the ten location problem. In Figure 6.5.6 we again see the first iteration taking a long time to complete for the subgradient method. Eventually, the subgradient method outperforms the border partitioning heuristic, however, it is only by a small amount.

6.6 Reinforcement Learning

A different direction we could take to find Nash equilibria in the stochastic game with multiple patrollers is to let agents learn the equilibria. In this section, we apply the different reinforcement learning algorithms which have been introduced in Chapter 2 to the multiple patroller problem.

We want to study how the different reinforcement learning algorithms converge to a Nash equilibrium. We consider two different metrics to measure the performance. First, we look at the Euclidean distance of the estimated value function $\tilde{V}$ from the true value function V, given by

$$\sqrt{\sum_{s \in \mathcal{S}} |\tilde{V}(s) - V(s)|^2}.$$

Note, we calculate the true value function exactly on these small examples using a linear program. As seen in Chapter 2, it is proven that the distance should converge to zero for the algorithms if the parameters of the algorithm meet certain assumptions.

However, the Euclidean distance does not show how the behavior of the patrollers is converging. For example, the patrollers could be playing a Nash equilibrium, but have their belief of the value function be off by a constant. If we only considered the distance

between the value functions, it would look like we had not converged to an equilibrium, when the strategies themselves have already converged. Therefore, we will also consider as our second metric the worst case expected reward that the patrollers can receive by playing their current strategy. Given a strategy $\boldsymbol{\pi}$ for the patrollers, the analysis of the previous chapter can be used to calculate the best response for the smugglers. If the patrollers are playing a strategy $\boldsymbol{\pi}^*$ which could form a Nash equilibrium, we have that the worst case value function v_{WC} is equal to the value function of the game $v_{WC} = v^*$. Otherwise, if the patrollers play a strategy which cannot be in a Nash equilibria, the worst case value function must be strictly worse than the value function $v_{WC} < v^*$.

We are using the same example of the border patrol model, which we will test the different algorithms on. The parameters for each model being solved are $k = 2$ patrollers, $n = 10$ locations, $r_i = 1$ and $C_i = 4$ for all $i \in [n]$, $m_{i,j} = (i - j)^2$ and $\gamma = 0.75$

For the tunable parameters of each algorithm, we have hand-picked ones which seem to provide the best possible convergence. Firstly, in the model based fictitious play algorithm we are using the parameters $\alpha_c = c^{-0.6}$, $\beta = c^{-0.8}$, and 10^6 iterations. Secondly, in the model free fictitious play algorithm we use $\alpha_c = c^{-0.5}$, $\beta = c^{-0.55}$, $\epsilon = 0.01$ and 10^7 iterations. Finally, in the Q-learning algorithm we use the parameters $\alpha_c = c^{-0.9}$, $\beta = c^{-1}$, $\tau = 5(1+0.5\log(c))^{-1}$ and 10^8 iterations. All the chosen parameters meet the respective assumptions for their algorithms to reach convergence.

Model Based Fictitious Play

We first consider the model based fictitious play algorithm by Sayin et al. (2022) which can be seen in Algorithm 1.

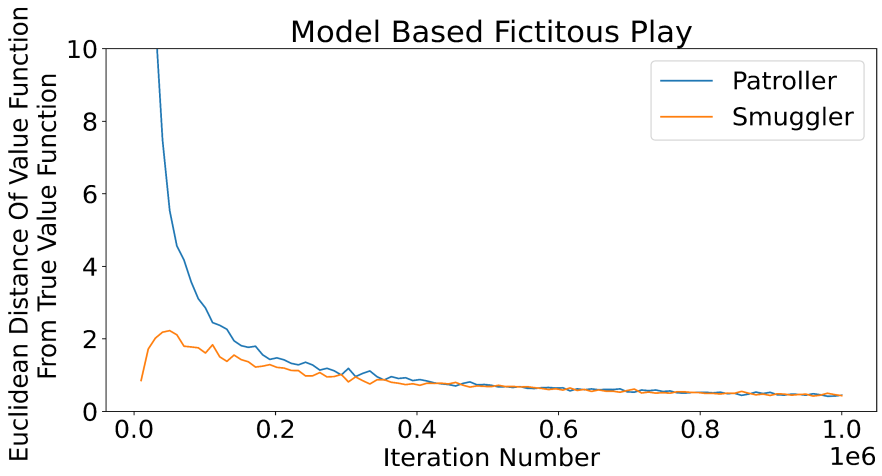


Figure 6.6.7: Error of value functions with model free fictitious play

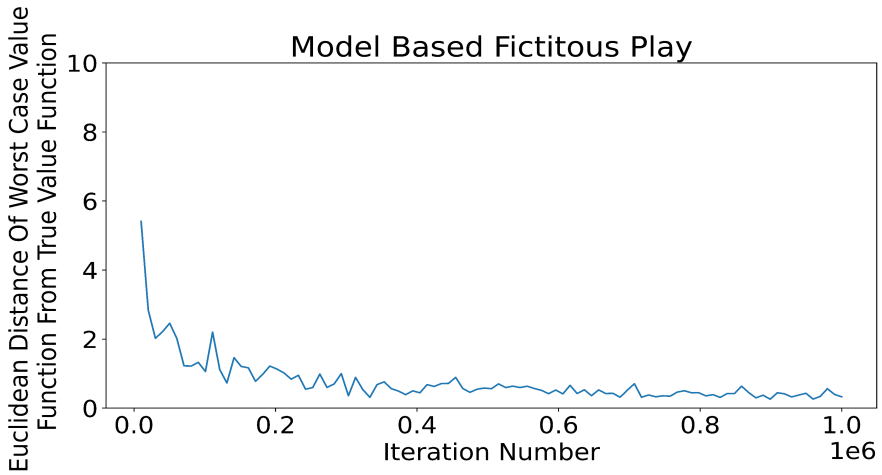


Figure 6.6.8: Worst case suboptimality of patroller's policy with model free fictitious play

We see fast and smooth convergence in the model based fictitious play setting, compared to the next two algorithms. These observations are expected since the agents now know fully the environment that they are in, and so have more information when taking their actions.

Model Free Fictitious Play

We next look at the model based fictitious play algorithm by Sayin et al. (2022) which can be seen in Algorithm 2.

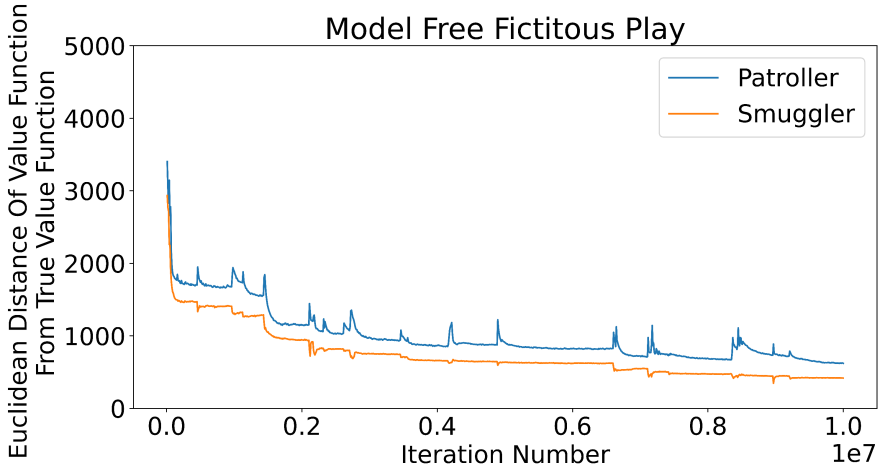


Figure 6.6.9: Error of value functions with model based fictitious play

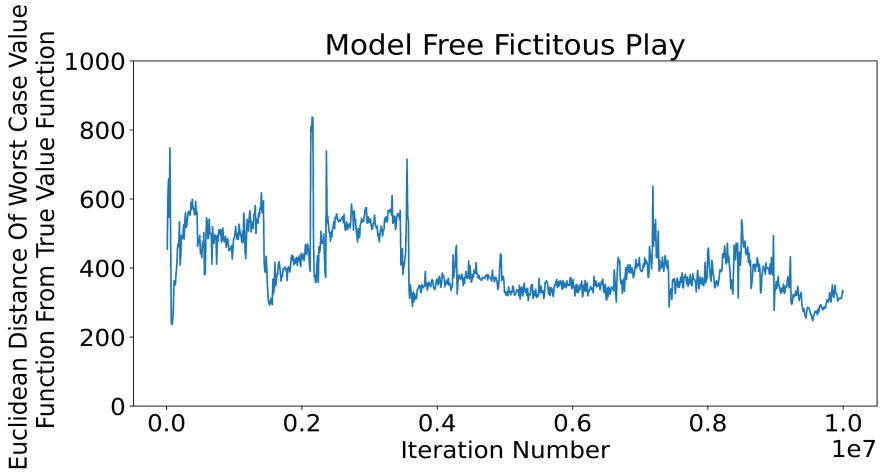


Figure 6.6.10: Worst case suboptimality of patroller's policy with model based fictitious play

We can see in the model free fictitious play setting, the convergence of the algorithm is erratic. There are large spikes where performance is lost under both metrics, it is an open question as to why and when these occur in more detail. Eventually, the algorithm does converge to an equilibrium, however, we will see that other algorithms do this with

fewer iterations. The interest of a model free algorithm though is we can see with no prior knowledge of the model, agents' behavior will converge to a Nash equilibrium.

Q-Learning Results

Finally, we apply the Q-learning algorithm by Sayin et al. (2021) which can be seen in Algorithm 3.

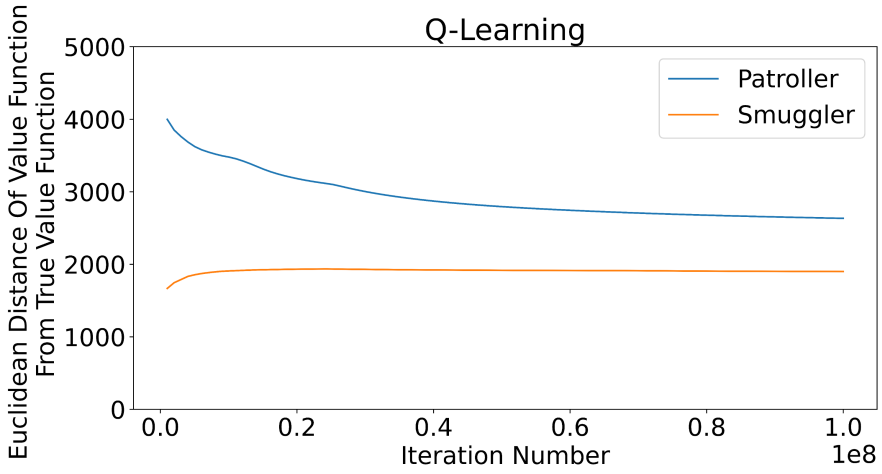


Figure 6.6.11: Error of value functions with Q-learning

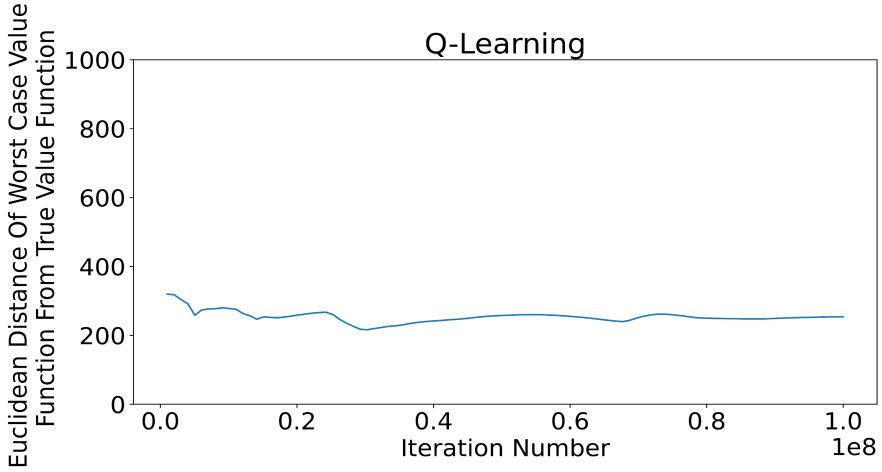


Figure 6.6.12: Worst case suboptimality of patroller's policy with Q-learning

We see that we get a slow rate of convergence with the Q-learning algorithm that we implemented. The problem is likely caused by the tunable parameters picked for the

algorithm. We were unable to find parameters which would give better performance, however, we did not do an exhaustive search through possible combinations.

6.7 Conclusion

In this chapter, we have introduced a multiple patroller extension to the framework of Chapter 5. We have considered three different approaches: exact methods, heuristics, and reinforcement learning. We have detailed how we would approach the problem using the method, and provided a numerical analysis of its performance.

There are multiple ways in which each of the approaches discussed in this chapter could be extended with future work.

Firstly, we could look to improve the convergence of the subgradient method that we have introduced. One way we could look at doing this would be to change the algorithm for subgradient descent. We have only considered one algorithm for doing so, and within the algorithm had limited time to optimize the hyperparameters of the algorithm. It is likely that the algorithm could be tuned to work better on our problem, or that a different algorithm exists within the literature to achieve better performance. Related to this problem, one idea we did not look at implementing yet is varying the length of the epochs within the subgradient method depending on the convergence of the value iteration algorithm. We hypothesize that having smaller epochs when the value iteration algorithm is far from convergence, and then increasing the size of the epoch as the subgradient method converges would achieve better performance. There are multiple ways which this could be implemented, and it would be an interesting direction for future research on the problem.

We only consider two heuristics in this chapter, and there is the open problem of finding other heuristics which would achieve a better performance in the model. Furthermore, for the two heuristics we have introduced there could be possible extensions to

improve them. Instead of considering a myopic policy for the patroller, maybe instead it would be possible to consider a heuristic which also takes into account the rewards obtained from the next time-step. Introducing an ability to look ahead would allow the smugglers to plan for future movements, instead of only trying to stop the smugglers in the immediate time-step. However, this would obviously increase the complexity of finding their strategy and so this is a direction for future work. When discussing the heuristic to partition the border, we only consider partitions which split the border into segments of equal size. Whilst this seems to be the intuitive partition for borders, we have no analytical proof that it is the best way to do so. An open question would be for different shapes and sizes of borders what the optimal partition of the border is.

When applying reinforcement learning to our problem, we have considered only two possible algorithms from the literature. There exist other reinforcement learning algorithms for stochastic games, and it would be an interesting area for future work to see if they perform better. Additionally, with the algorithms which we did implement, we did not have time to fully optimize their performance on our problem. There is room for improvement with further work on the problem to pick better hyperparameters to achieve faster convergence to the Nash equilibrium.

A final open problem for the chapter would be to consider if particular classes of problem instances would allow for easier analysis. For example, if only line or circle borders were considered instead of a general border, we could perhaps prove additional properties about the game. We would like to look at questions such as if the patrollers would ever cross over, and how they space themselves apart. However, we cannot analyse these questions within a general setting and so a simplification of the model would be necessary.

6.8 Appendix

Algorithm 10: Euclidean projection of a vector onto the probability simplex (Wang and Carreira-Perpinán, 2013)

Input: $\boldsymbol{y} \in \mathbb{R}^N$

1 Sort the vector $\boldsymbol{y}$ into a new vector $\boldsymbol{u}$ such that it is in descending order, i.e.

$$u_1 \geq u_2 \geq \cdots \geq u_N$$

2 Find,

$$\rho = \max \left\{ j \in [N] \;\middle|\; u_j + \frac{1}{j} \left(1 - \sum_{i=1}^{j} u_i \right) > 0 \right\}$$

3 Define,

$$\lambda = \frac{1}{\rho} \left(1 - \sum_{i=1}^{\rho} u_i \right)$$

Output: $\boldsymbol{x}$ such that $x_i = \max\{y_i + \lambda, 0\} \; \forall \; i \in [N]$

Chapter 7

Conclusions

We conclude the thesis by summarizing the contributions made and giving directions for future research on the topic.

7.1 Contributions

In Chapter 4 we presented a novel framework to consider the communication and cooperation between smugglers in a border patrol problem. We first detailed how equilibria are defined under each setting, and then prove analytical results which hold across them all. For each specific case, we then delved deeper and proved which strategies are equilibria.

Chapter 5 examined a new stochastic game model for patrolling a border. We have extended existing models in the literature by considering a continuous action space for the smugglers, which complicates the analysis of the game. We proved a number of properties of equilibria in the game, and presented algorithms to compute them.

Finally, Chapter 6 contributed interesting directions to consider the multiple patroller extension of Chapter 5. We looked at methods to find the equilibria exactly, heuristics, and reinforcement learning approaches to the problem.

7.2 Further work

To extend the work of Chapter 4 there could be a number of interesting directions taken. The model could be developed further by adding in multiple patrollers. Currently, the analysis of the chapter is currently limited to only a single patroller defending against multiple smugglers. However, it would be a more realistic assumption that we could have a number of patrollers patrolling the border. The analysis would be complicated, but it is an attractive direction for future research in the model. A different extension which could be looked into is introducing multiple item types which are available to smuggle. The model currently only accounts for a single type of contraband to be trafficked by the smugglers, however, in a real-life scenario there could be many. An example of particular interest could be for the smuggling of illicit drugs, the different types of drugs and how the market for each of them depends differently on the supply. Finally, we would like to consider the game over multiple time periods, as we have for the models in Chapters 5 and 6. The most natural factor to incorporate would be movement costs for the patroller, and to see how the communication and cooperation between smugglers could affect this. However, the model quickly becomes too complex to either computationally or analytically find solutions.

There are multiple extensions which could be looked at to extend the work of Chapter 5. Firstly, multiple patrollers can be added, which we have looked into with Chapter 6. Secondly, we could consider the patroller having the ability to catch multiple smugglers in the game. A camera or drone could be able to surveil a larger area than one smuggler could attack, and therefore, multiple smugglers might get caught at once. Thirdly, we could consider further constraints on the quantities of items which the smugglers can send to the border. An intuitive extension would be to allow smugglers to stockpile the items which they do not send, and then in a future turn make a larger attack. Such an extension would complicate the analysis of the game, for example we would no longer have a single controller stochastic game, but might reveal interesting

behavior.

One extension possible to Chapter 5 is the inclusion of multiple patrollers, which we have looked at in more detail with Chapter 6. However, there are a number of ways Chapter 6 could be developed with future research. Firstly, the subgradient descent approach to finding equilibria in the model could be improved upon to speed up the computational time. We could consider a more intelligent starting point for the algorithm, for example, by using one of the heuristics discussed later in the Chapter. Another interesting adaptation would be to look at having the number of epochs with which the subgradient method runs for to depend on the accuracy of the value function. If the state values are still inaccurate, it seems intuitive that less time should be spent finding the exact solution to the subgradient method. Secondly, there could be more heuristics developed for the game. We only discuss two possible heuristics in Chapter 6, and there could be a range of interesting and potentially more accurate heuristics available, which we have not discussed in this thesis. Finally, further research could be taken with reinforcement learning approaches to the multiple patroller problem. A direction of particular interest would be to consider if the smuggler aggregation property of the game could be utilized to improve the rate of convergence when learning equilibria.

Bibliography

Alpern, S., Bui, T., Lidbetter, T., and Papadaki, K. (2022a). Continuous patrolling games. *Operations Research*, 70(6):3076–3089.

Alpern, S., Chleboun, P., Katsikas, S., and Lin, K. Y. (2022b). Adversarial patrolling in a uniform. *Operations Research*, 70(1):129–140.

Alpern, S. and Gal, S. (2006). *The theory of search games and rendezvous*, volume 55. Springer Science & Business Media.

Alpern, S., Lidbetter, T., and Papadaki, K. (2019). Optimizing periodic patrols against short attacks on the line and other networks. *European Journal of Operational Research*, 273(3):1065–1073.

Alpern, S., Morton, A., and Papadaki, K. (2011). Patrolling games. *Operations Research*, 59(5):1246–1257.

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica: Journal of the Econometric Society*, 55(1):1–18.

Baniya, S. (2023). Europe key to Syrian regime's amphetamine trade, finds report. https://www.euronews.com/2023/09/24/europe-key-to-syrian-regimes-amphetamine-trade-finds-report. Accessed: 17-01-2024.

Baston, V. J. and Bostock, F. (1991). A generalized inspection game. *Naval Research Logistics (NRL)*, 38(2):171–182.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of political economy*, 76(2):169–217.

Berger, U. (2005). Fictitious play in 2×n games. *Journal of Economic Theory*, 120(2):139–154.

Bichler, G., Malm, A., and Cooper, T. (2017). Drug supply networks: a systematic review of the organizational structure of illicit drug trade. *Crime Science*, 6(1):1–23.

Bier, V., Oliveros, S., and Samuelson, L. (2007). Choosing what to protect: strategic defensive allocation against an unknown attacker. *Journal of Public Economic Theory*, 9(4):563–587.

Brown, G., Carlyle, M., Salmerón, J., and Wood, K. (2006). Defending critical infrastructure. *Interfaces*, 36(6):530–544.

Brown, G. W. (1951). Iterative solution of games by fictitious play. *Act. Anal. Prod Allocation*, 13(1):374.

Chalmers, J., Bradford, D., Jones, C., et al. (2009). How do methamphetamine users respond to changes in methamphetamine price. *Crime and Justice Bulletin*, 134:1–16.

Chvátal, V. (1983). *Linear programming*. Macmillan.

Danskin, J. M. (1967). *The theory of max-min and its application to weapons allocation problems*. Springer-Verlag, Berlin; New York.

Darlington, M., Glazebrook, K. D., Leslie, D. S., Shone, R., and Szechtman, R. (2023). A stochastic game framework for patrolling a border. *European Journal of Operational Research*, 311(3):1146–1158.

Filar, J. (1985). Player aggregation in the traveling inspector model. *IEEE Transactions on Automatic Control*, 30:723–729.

Filar, J. and Schultz, T. (1986). The traveling inspector model. *OR Spectrum.*, 8(1):33–36.

Filar, J. and Vrieze, K. (2012). *Competitive Markov decision processes*. Springer Science & Business Media.

Fox, B. (1966). Discrete optimization via marginal analysis. *Management Science*, 13(3):210–216.

Freedman, E. (2022). Man pleads guilty to smuggling $739k of wildlife, including crocodiles. https://www.independent.co.uk/climate-change/news/animal-smuggling-lizards-crocodiles-b2152959.html. Accessed: 17-01-2024.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*, volume 2. MIT press.

Garnaev, A. Y. (1994). A remark on the customs and smuggler game. *Naval Research Logistics (NRL)*, 41(2):287–293.

Goodman, J. (2021). Takeaways from AP and Univision China fishing investigation. https://apnews.com/article/china-fish-south-america-pacific-ocean-overfishing-733633c3d67c1f8322b382fd9ec10183. Accessed: 02-02-2023.

Grant, J. A., Leslie, D. S., Glazebrook, K., Szechtman, R., and Letchford, A. N. (2020). Adaptive policies for perimeter surveillance problems. *European Journal of Operational Research*, 283(1):265–278.

Gutierrez, G. and Henkel, A. (2021). Fentanyl seizures at U.S. southern border rise dramatically. https://www.nbcnews.com/politics/immigration/fentanyl-seizures-u-s-southern-border-rise-dramatically-n1272676. Accessed: 02-02-2023.

Hazan, E. and Kakade, S. (2019). Revisiting the Polyak step size. *arXiv preprint arXiv:1905.00313*.

Hochbaum, D. S. (1994). Lower and upper bounds for the allocation problem and other nonlinear optimization problems. *Mathematics of Operations Research*, 19(2):390–409.

Hofbauer, J., Sigmund, K., et al. (1998). *Evolutionary games and population dynamics*. Cambridge university press.

Kaplan, H., Kozma, L., Zamir, O., and Zwick, U. (2019). Selection from heaps, row-sorted matrices, and X + Y using soft heaps . *Symposium on Simplicity in Algorithms*, page 5:1–5:21.

Kikuta, K. and Ruckle, W. H. (2002). Continuous accumulation games on discrete locations. *Naval Research Logistics (NRL)*, 49(1):60–77.

Kleiman, M. and Kilmer, B. (2009). The dynamics of deterrence. *Proceedings of the National Academy of Sciences*, 106(34):14230–14235.

Lin, K., Atkinson, M., Chung, T., and Glazebrook, K. (2013). A graph patrol problem with random attack times. *Operations Research*, 61:694–710.

Lin, K., Atkinson, M., and Glazebrook, K. (2014). Optimal patrol to uncover threats in time when detection is imperfect. *Naval Research Logistics (NRL)*, 61:557–576.

Lin, K. Y. (2022). Optimal patrol of a perimeter. *Operations Research*, 70(5):2860–2866.

Lindelauf, R., Borm, P., and Hamers, H. (2009). The influence of secrecy on the communication structure of covert networks. *Social Networks*, 31(2):126–137.

Maitra, A. and Parthasarathy, T. (1970). On stochastic games. *Journal of Optimization Theory and Applications*, 5(4):289–300.

McGrath, R. G. and Lin, K. Y. (2017). Robust patrol strategies against attacks at dispersed heterogeneous locations. *International Journal of Operational Research*, 30(3):340–359.

Monderer, D. and Shapley, L. S. (1996). Potential games. *Games and economic behavior*, 14(1):124–143.

Nash, J. (1951). Non-cooperative games. *Annals of Mathematics*, 54(2):286–295.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.

Norton, D. A. G. (1988). On the economic theory of smuggling. *Economica*, 55(217):107–118.

Ojewale, O. (2021). State collusion perpetuates oil smuggling across Nigeria-Cameroon borders. https://enactafrica.org/enact-observer/state-collusion-perpetuates-oil-smuggling-across-nigeria-cameroon-borders. Accessed: 02-02-2023.

Papadaki, K., Alpern, S., Lidbetter, T., and Morton, A. (2016). Patrolling a border. *Operations Research*, 64:1256–1269.

Pita, J., Jain, M., Marecki, J., Ordóñez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., and Kraus, S. (2008). Deployed armor protection: the application of a game theoretic model for security at the Los Angeles International Airport. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems: industrial track*, pages 125–132.

Politi, A. (1997). *European Security: The new transnational risks*. Institute for Security Studies, Western European Union.

Raghavan, T. (2003). Finite-step algorithms for single-controller and perfect information stochastic games. In *Stochastic games and applications*, pages 227–251. Springer.

Rhodes, W., Johnston, P., Han, S., McMullen, Q., and Hozik, L. (2000). *Illicit drugs: Price elasticity of demand and supply*. Abt Associates.

Richard, L. (1972). Urban police patrol analysis. *MIT Press, Cambridge, MA*.

Robinson, J. (1951). An iterative method of solving a game. *Annals of mathematics*, pages 296–301.

Ruan, S., Meirina, C., Yu, F., Pattipati, K. R., and Popp, R. L. (2005). Patrolling in a stochastic environment. In *10th Intl. Command and Control Research and Tech. Symp.*

Ruckle, W. (2001). Accumulation games. *Sci. Math. Japan*, 54(1):173–203.

Sack, J.-R. and Urrutia, J. (1999). *Handbook of computational geometry*. Elsevier.

Sah, R. K. (1991). The effects of child mortality changes on fertility choice and parental welfare. *Journal of Political Economy*, 99(3):582–606.

Savage, C. and Bergman, R. (2023). U.S. seized Iranian oil over smuggling incident that escalated tensions in gulf. https://www.nytimes.com/2023/09/06/us/politics/iran-oil-sanctions-violations.html. Accessed: 17-01-2024.

Sayin, M., Zhang, K., Leslie, D., Basar, T., and Ozdaglar, A. (2021). Decentralized Q-learning in zero-sum Markov games. *Advances in Neural Information Processing Systems*, 34:18320–18334.

Sayin, M. O., Parise, F., and Ozdaglar, A. (2022). Fictitious play in zero-sum stochastic games. *SIAM Journal on Control and Optimization*, 60(4):2095–2114.

Shapley, L. (1964). Some topics in two-person games. *Advances in game theory*, 52:1–29.

Shapley, L. S. (1953). Stochastic games. *Proceedings of the National Academy of Sciences*, 39(10):1095–1100.

Sheikh, M. A. (1974). Smuggling, production and welfare. *Journal of International Economics*, 4(4):355–364.

Shieh, E., An, B., Yang, R., Tambe, M., Baldwin, C., DiRenzo, J., Maule, B., and Meyer, G. (2012). Protect: A deployed game theoretic system to protect the ports of the United States. In *Proceedings of the 11th international conference on autonomous agents and multiagent systems-volume 1*, pages 13–20.

Smith, J. C. and Song, Y. (2020). A survey of network interdiction models and algorithms. *European Journal of Operational Research*, 283(3):797–811.

Stewart, M. G., Ellingwood, B. R., and Mueller, J. (2011). Homeland security: A case study in risk aversion for public decision-making. *International Journal of Risk Assessment and Management*, 15(5-6):367–386.

Stone, L. D. (1976). *Theory of optimal search*. Elsevier.

Sutton, R. S. and Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

v. Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320.

Vrieze, O. and Tijs, S. (1982). Fictitious play applied to sequences of games and discounted stochastic games. *International Journal of Game Theory*, 11:71–85.

Wang, W. and Carreira-Perpinán, M. A. (2013). Projection onto the probability simplex: An efficient algorithm with a simple proof, and an application. *arXiv preprint arXiv:1309.1541*.

Washburn, A. and Wood, K. (1995). Two-person zero-sum games for network interdiction. *Operations research*, 43(2):243–251.

Wickramasekera, N., Wright, J., Elsey, H., Murray, J., and Tubeuf, S. (2015). Cost of crime: A systematic review. *Journal of Criminal Justice*, 43(3):218–228.

Yang, R., Ford, B. J., Tambe, M., and Lemieux, A. (2014). Adaptive resource allocation for wildlife protection against illegal poachers. In *Aamas*, pages 453–460.