City Research Online

City St George's, University of London

Citation: Scaramangas, A. P. (2023). Evolutionary and Eco-Evolutionary Stability in Aposematic Prey Populations. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/30329/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Evolutionary and Eco-Evolutionary Stability in Aposematic Prey Populations

Alan Philip Scaramangas

City, University of London

Department of Mathematics

Thesis submitted for the degree of

Doctor of Philosophy

March 2023

Contents

List of Figures **iii**

List of Tables **iii**

Acknowledgements **v**

Declaration of Authenticity **vi**

Abstract **vii**

Preface **viii**

1 Game Theory and Evolution **1**

1.1 Introductory remarks 1

1.2 Concepts from classical and evolutionary games 4

1.3 Static concepts in evolutionary game theory 9

2 Aposematic Defence **16**

2.1 Background 16

2.2 Model description 21

2.3 The evolutionary stability of aposematism 30

2.4 A marginal fitness interpretation of evolutionary stability 47

3 An exploration into the ESS continuum **51**

3.1 The monotonicity of the ESS continuum 52

3.2 An explicit procedure for determining local ESSs 55

3.3 Explicit examples of evolutionarily stable outcomes: A numerical analysis 62

3.4 Remarks on the convergence stability of local ESSs 67

3.5 Some discussion on the honest signalling of aposematism 69

4 Theoretical and genetic algorithm predictions on ESSs with non-zero rates of background mortality **74**

4.1 ESS analysis for the genetic algorithm model 75

4.2 A description of the simulation 83

4.3 Solutions with vanishing background mortality 86

4.4 Solutions with non-zero background mortality $\boldsymbol{\lambda > 0}$ 91

4.5 Discussion . . . 95
4.6 Code . . . 98

5 Co-existence & mimicry: a first approach **104**
5.1 Two-type co-existence . . . 105
5.2 Models vs. mimics . . . 120
5.3 A stable point solution . . . 123
5.4 A stable continuum of solutions . . . 131

6 Batesian Mimicry **140**
6.1 Background into Batesian mimicry . . . 140
6.2 A sampling theory approach . . . 143
6.3 A probability theory approach . . . 148
6.4 Ecological stability . . . 154
6.5 Evolutionary stability . . . 155
6.6 Eco-evolutionary stability and the confluent hypergeometric function . . . 162
6.7 An example of eco-evolutionarily stable Batesian mimicry . . . 165
6.8 Items for discussion . . . 174
Bibliography . . . 179

List of Figures

3.1 Stable and unstable cluster sizes 60
3.2 Stability bands by cluster size 61
3.3 The intersection method for finding equilibria 64
3.4 Equilibrium curves under varying pressures of predation 65
3.5 ESS curves for different regimes of prey sensitivity 67

4.1 Example functions 76
4.2 ESS for vanishing background mortality and vanishing local relatedness #1 86
4.3 ESS for vanishing background mortality and vanishing local relatedness #2 89
4.4 ESS for vanishing background mortality and non-vanishing local relatedness 91
4.5 ESS for non-vanishing background predation and vanishing local relatedness 92
4.6 Varying background mortality for vanishing local relatedness 93
4.7 Varying background mortality with non-vanishing local relatedness 95

5.1 Illustrating ecological stability 110
5.2 Stable automimicry continuum 138

6.1 Beta-distributed density functions with various shape parameters 152
6.2 The proportion of mimics as a beta-distributed density function (with incremental variation in the mean) 172
6.3 A stable Batesian mimicry continuum 172

List of Tables

2.1 The parameters and functions of the model. 26

6.1 Point-by-point checking of ecological and evolutionary stability conditions on continuum . . . 173

Acknowledgements

I am indebted to Prof. Mark Broom for believing in me and for supporting me during these (now almost) five intense years of academic study. I feel honoured to have been introduced to the world of evolutionary games by a true and leading expert who also happens to be a very nice and humorous person. During this time we have established a good working relationship, which lead to two publications (and to the possibility of a third, which we may pursue in the near future - see chapter 6). This research would not have been possible without the generous *EPSRC* stipend and I am especially grateful to Mark for considering me as a candidate for the awarding of this studentship. I am also grateful to Prof. Graeme Ruxton of St. Andrew's University for a number of reasons. Graeme read through our first publication (on which chapter 3 is based), co-authored the second (on which chapter 4 is based) and read through and provided comments on the last chapter of the thesis. Graeme also introduced us to a talented postgraduate student, Ms. Anna Rouvière who kindly ran a large number of simulations a selection of which are shown in chapter 4. I am very grateful to my parents, my brother and my (now ex-) girlfriend for their enormous patience and support. I have truly enjoyed my time at City University and have made a number of good friends during this period; after the pandemic subsided I found that coming into the department on a daily basis and interacting with my peers in person made long working days seem like a breeze. I am very thankful to everyone in this academic community.

Declaration of authenticity

This report contains genuine work conducted originally by the author. The work presented herein has not been submitted and/or accepted for the award of any other degree or diploma in any university. To the best of my knowledge and belief, this thesis contains no materials previously published or written by other person, except where due references has been made.

Location: London, United Kingdom

Date: 10/3/2023

Name: Alan Philip Scaramangas

Abstract

The term aposematism or "warning colouration" (as first characterised by Alfred Russel Wallace in 1877) describes the process by which defended organisms (animals or plants) advertise their unprofitability to potential predators to gain selective advantage. The first part of the thesis explores the relationship between evolutionarily stable levels of signalling and defence within the context of a game-theoretical, prey-predator setup in which the prey population consists of a single type. While it is implicitly assumed that the prey population is large enough to be considered effectively infinite the evolution of prey traits is also explored for intermediate-sized populations within the context of genetic algorithm approach. In the later chapters considerable effort is devoted to extending the mentioned predator-prey description to systems in which the prey population consists of two types, including a model and a mimic. This modification leads us naturally into the celebrated adaptive mechanism named after Henry Walter Bates, Batesian mimicry. In Batesian mimicry complexes individuals from a palatable (mimic) species resemble individuals from an unpalatable (model) species to gain protection against predators. While there is ample empirical evidence to suggest that individuals from one species may gain selective advantage by resembling individuals from another, the mathematical modelling of Batesian mimicry is rather limited. We predict that models and mimics can co-exist along a continuum of solutions (representing the conspicuousness, noxiousness, and average mimic-to-model proportion) that are both ecologically and locally evolutionarily stable. We establish a number of novel results that confirm both common sense intuition and a considerable body of related works.

Preface

In this thesis we provide a mathematical description of aposematism (the conspicuous signalling of secondary defences) by building on and extending an existing game-theoretical model of the process. The thesis consists of two major parts with individual chapters as follows: the first chapter is introductory; the second, third and fourth deal with aposematic prey populations of a single species and in chapters five and six we deal with populations that consist of two types (of either the same or of a different species). In the last chapter we provide a novel extension of the model to systems of Batesian mimicry, which is notable considering the lack of mathematical modelling of this phenomenon.

In the first chapter we provide a brief historical overview of the theory of classical and evolutionary games; we introduce notions of *strategy* (*pure* or *mixed*), and *payoff*, and explain how the assumption of evolution by natural selection renders these notions appropriate in the context of biology. Of critical importance is the notion of an *Evolutionarily Stable Strategy* (ESS), which we define and explore through a number of theorems and worked examples (see *Hawk-Dove* and more notably *The War of Attrition* in which we recover a continuum of solutions that are similar to those seen in chapter 3). Indeed, while the War of Attrition is a linear game that is considerably simpler than the (non-linear) game that we explore in this thesis the mentioned ESS continuum is a remarkable common feature. The reader familiar with these notions may skip this first chapter entirely, although perhaps not the ESS analysis of the War of Attrition provided at the end of the chapter.

In the second chapter we introduce aposematism in the context of predator-interaction, following (mostly) the layout of the original publication. The prey population is initially treated as consisting of a single-species (although see chapters 5 and 6), wherein each individual is distinguished by means of two continuous and independently varying traits (signalling and defence). There are a number of differences between the presentation in this chapter and the original. For instance, we demonstrate how the known expressions for perceived aversiveness may be recovered in the limit as the overall proportion of mutants vanishes, which is insightful, particularly from then modelling perspective. Previous readership has found the relationship between the overall proportion of mutants and the role of local relatedness in influencing perceived aversiveness confusing. To that, we provide detailed explanations of how a territorial division of the habitat among predators can allow for a cluster of identical (focal) individuals to pose threat of invasion without contributing significantly to the overall proportion of prey. We include detailed verbal explanations of the conditions for ESS by means of a marginal fitness interpretation, which has, perhaps, previously been lacking. In addition, we make a number of clarifications relating to the properties of the functional forms involved, which leads us smoothly onto chapter 3.

In the third chapter we argue that the original model can be extricated from two restricting conjectures that were made upon its publication and hence demonstrate that it can account for a larger plethora of physical systems than had previously been assumed. The chapter is based on a publication co-authored by myself and Prof. Mark Broom in 2022 in which we argue that evolutionarily stable levels of defence need not be positively related with the conspicuousness and further that for given level of signal strength may correspond more than one optimal levels of defence. Through example we demonstrate that it is not generally possible to arrive at an explicit relationship between evolutionarily stable levels of signalling and the defence being signalled and argue that in such a general class of examples the *Implicit Function Theorem* (in $\mathbb{R}^2$) is required to describe the monotonicity of the continuum. The honest signalling of prey defences (i.e. whether more conspicuous morphs are better defended) remains puzzling not only among empiricists but also among

those seeking to model aposematism. Our deduction that the mentioned model can account for both honest and dishonest signalling (within the context of evolutionary stability) is therefore an important one.

In the fourth chapter we implement a genetic algorithm approach to determine the evolution of prey traits in a finite (intermediate-sized) population of prey subject to small random mutations. The chapter is based on the paper (status: submitted) co-authored by myself, Prof. Mark Broom, Prof. Graeme Ruxton and Ms. Anna Rouvière in 2022 - we are indebted to Ms. Rouvière for the countless runs of requested simulations using the statistical package *R*. In this we explore the possibility of evolutionarily stable outcomes in which predation is not the only source of death, such that the parameter describing the rate of background mortality is in some cases assumed to be non-vanishing. Previous attempts (including those of chapters 3,5 and 6) have taken the mentioned parameter to be zero. Among the various different insights of this chapter we predict (and confirm via simulation) the perhaps intuitive result that the equilibrium level of defence shrinks with increasing levels of background mortality. We also predict that while marginal differences in the mutant fitness is the stronger driver of prey trajectories in intermediate-sized populations the absolute resident fitness is also important, especially when local relatedness is considerable.

The notion of bi-stability is central both to chapter 5 and to chapter 6 and describes a situation in which two types co-exist over the short-term (ecological) time-scale and over the prolonged (evolutionary) time-scale such that the latter relies on the former holding true. We should also remark that evolutionary stability is in this context understood as the requirement that each type is evolutionarily stable against mutations that are local to its type. While the focus of chapter 5 is to model co-existence in the context of mimicry the presentation is initially kept general and indeed a number of interesting conclusions are drawn that do not strictly relate to mimicry. In particular, it is observed that the co-existence of: (i) two attractive types in which at least one has non-zero signalling component is not evolutionarily stable - this also implies that the mimicry complexes involving two conspicuous and attractive types is unstable; (ii) one aversive type and one attractive type is not evolutionarily stable if the attractive type has the stronger signalling component; (iii) two aversive types with vastly different signalling components is not ecologically stable. In the second part of this chapter we narrow our attention to instances of co-existence that are strictly mimetic and hence recover solutions that are manifest as a continuum over the extended (to account for mimetic load) strategy space. While throughout the thesis mutation is facilitated by means of local clustering the specific implementation of the relatedness parameter in this chapter renders the later discussion more adept to the modelling of (automimetic/Browerian) mimicry complexes in which the two types belong to the same species.

The sixth chapter deals explicitly with mimicry complexes and while the intention is still to determine outcomes that are bi-stable the approach to modelling co-existence in this varies considerably in comparison with chapter 5. For instance, while previously the presence of the second type is accounted for using a real variable that measures the background concentration of mimics this is now treated as a continuous (beta-distributed) random variable such that the proportion of models to mimics is naturally taken to fluctuate about some mean value throughout the habitat. In addition, even though mutation is still facilitated through local clustering the presence of, say type-1 (model) mutants in some locality does not influence the proportion of models to mimics in that site. That is, the operational definition of the relatedness parameter is in this chapter understood as the concentration of focal types taken as a proportion over the number of individuals of that type and not as a proportion over the total number of individuals in the site (chapter 5). This is a key distinguishing feature that renders the approach of this chapter better-suited to modelling Batesian mimicry complexes in which models and mimics belong to different species. Much like in the previous chapter through an explicit worked example we recover bi-stable solutions manifest as a continuum in the extended strategy

space and observe an increasing relationship between the mimetic load and the signal strength associated with the complex.

Readers familiar with the theory of games and evolutionary games may skip the first chapter completely; that said, perhaps a revisiting of the ESS analysis on the War of Attrition is a worthwhile investment. Since exact knowledge of the mathematical modelling of aposematism cannot be assumed we encourage readers to go through chapter 2 before focusing on any specific chapter thereafter. Readers interested in the mathematical modelling of aposematism in prey populations that consist of a single type are thus encouraged to read chapters 2, 3 and 4. Readers interested in the applications of aposematism to genetic algorithms should read chapters 2 and 4. Readers interested in the automimicry and or Batesian mimicry systems are encouraged to read chapters 2,5 and 6 (and if time is of essence perhaps chapters 2 and 6 only).

Chapter 1

Game Theory and Evolution

The purpose of this chapter is to motivate the connection between mathematics and the life sciences. It begins with a historical account of how classical and evolutionary games first developed, mentioning some of the main contributors in these areas and how some of the key concepts have branched out into sub-fields of evolutionary games that remain active at present. A distinction is drawn between the static and the dynamical approach in evolutionary games, differing primarily on the timescales that these consider and in the techniques that are implemented for analysis. The focus of the report is on the former and in closing some useful definitions and results from game theory are provided (with proofs omitted but with references to these indicated suitably) and two characteristic examples are analysed (namely the Hawk-Dove game and the War of Attrition).

1.1 Introductory remarks

The connection between the life sciences and mathematics is not been an obvious one and has grown more strong over the past decade than over any other period in the past. Arguably, the earliest developments in mathematical biology began during the early 20^{th} century with the birth of population genetics (the study of changes in allele frequencies within single-species populations) with characteristic works comprising Hardy (1908), Weinberg (1908), Wright (1930) and Fisher (1930) among others, and the subsequent study of population dynamics in predator-prey systems (see for example Lotka, 1925 and Volterra, 1926). In this report we consider a model in the context of evolutionary game theory, a branch which developed out of these earlier works in mathematical biology but which borrowed many of the fundamentals from game theory.

The emergence of game theory

Games have been played by people for thousands of years and within these players have devised different methods for optimising their play to increase their chances of winning (see Broom and Rychtár, 2013). There exist (and have existed) countless numbers of games, some involving perfect information (in which players have complete knowledge of the state of the game at every turn - examples of these are board games, such as chess) and others involving imperfect information (typical examples being card games, in many of which the hand of the other player is unknown and can only be guessed). A game of chess involves a finite number of moves and in principle the optimal choice for each of these can be established (a process which is now configured using chess computers/programs). In 1913 it was proposed by Ernst Zermelo (and proved in 1928

by Laszlo Kalmár - see Zermelo, 1913 and Kalmár, 1928) that in any two-person zero-sum game with perfect information one or the other player can force a win (namely have a winning strategy) or the game will end in a draw. But is a solution to a game (such as the above) guaranteed to exist, can the process of finding an optimal strategy be generalised to study any game, and more furthermore, is there a framework within which games can be studied in a systematic fashion?

Optimal play is dependent on the choices that other players in the game make and determining this involves choosing the best out of a number of choices, whose outcomes are uncertain. Uncertainty and random events have admitted a challenge to mathematics outside of games [1] and (roughly speaking) admit the object of study of probability (whose foundations were first laid out in the form they have now during the 17^{th} century). Related problems have been (more narrowly) addressed within the realm of optimisation and (normative) decision theory (developed during the 20^{th} century), in which an optimal decision is sought subject to number of (known) constraints. This is quite similar to the subject of game theory, but where the latter has the added complexity that instead, these constraints are imposed by the remaining players of the game whose behaviour is uncertain (yet assumed to be rational) - see Broom and Rychtár (2013). In fact, the games to which Zermelo's theorem applies (an example of which is chess) are ones in which (assuming players are rational and have full knowledge of all past moves made) there exists an optimal move at every round and can therefore be solved in a similar way as optimisation problems of decision theory. In general however, problems in game theory exhibit uncertainty (as opponent strategies are unknown) making the question of how a game is solved very relevant and the answer to this not an obvious one.

Some could argue that game theory started in the 1713 when James Waldegrave provided a solution to a two-player game of cards known as *le Her*, although this feat was isolated and at the time considered an achievement of probability theory. The theory of games in the form that this now exists began in the 1920s, when Emil Borel introduced the minimax solution to two-player games. This was subsequently proved in von Neumann (1928), who later (in 1944) co-authored and published (together with Oskar Morgenstern) the book *Theory of Games and Economic behavior* - see von Neumann and Morgenstern (1944) - which is to this day considered one of the classics on the topic. A characteristic example of a two-player, non-cooperative game is that of the Prisoner's Dilemma, which originates in the 1950s when it formed part of an experiment conducted as part of a non-profit research initiative. Its game-theoretical formalisation is credited to Albert Tucker in the 1980s and has drawn increasing attention since its emergence - see Tucker (1980).

Another important development in the theory of (non-cooperative) games (especially so for the purposes of this report) is the concept of a Nash Equilibrium (N.E.), which was introduced by John Forbes Nash Jr. in the early 1950s - see Nash (1950) and Nash (1951) - and provides an answer to the question of finding the "best strategy". Furthermore, it provided footing for the various branches of game theory that emerged subsequently (including that of evolutionary game theory) and drew the important distinction between cooperative and non-cooperative games. Nash together with Reinhard Selten and John Harsanyi received the Nobel prize in economics (1994), where the latter two not only provided refinements to Nash's original N.E. concept but also contributed otherwise (and substantially) to economics and game theory. Selten introduced the notion of sub-game perfect equilibria (see Selten, 1965), which he used to study dynamic, extensive form games (cases of this included competition with only a few sellers) and the "trembling hand perfect equilibrium" for perturbed games in Selten (1975). During the 1960s Harsanyi introduced the notion of equilibrium selection - see Harsanyi (1966), addressing the issue of how one equilibrium can be chosen

[1] That said, it is true that some of the first challenges in probability had been inspired by games and gambling problems - see for example Broom and Rychtar (2013) for the insight provided by the Mersenne salon (including Fermat and Pascal) into the solution to gambling problems given in response to a challenge set by Chevalier de Mere.

over an other and proposed methods of analysing games with incomplete information (known as Bayesian games).

The emergence of evolutionary game theory

Up until this point we have discussed games that are played by rational individuals that seek to optimise their playing strategy. The rationality assumption is so fundamental to game analysis that it would appear strange for its principles to be applicable in the biological setting. Animals cannot be assumed to behave rationally, and further, the notion of "reward", "strategy", or even of a "game" for that matter, would have to be relaxed substantially to cater for interanimal interaction. Interestingly, however, the connection is a strong one, with Darwin's theory of evolution by natural selection providing firm footing (see Darwin, 1859). In particular, rationality may be replaced by natural selection, reward to an individual may be replaced by its fitness (Darwinian - usually measured in terms of the number of offspring produced) and with a strategy describing the genetic programming of a certain animal individual (subsequently passed onto its offspring). This setup thus describes (usually large) populations of individuals playing different strategies and engaging in contests (examples of which are commonly drawn from classical game theory) so as to maximise their own fitness (once more, see Broom and Rychtár, 2013 for many examples of contests). By definition, the fittest strategies will prevail over others (more offspring) and the compositions will change over successive generations. Broadly speaking, this is the subject of evolutionary game theory, although depending on the time-scale of the process considered, the nature of the analysis may vary quite noticeably (see a comparison of the static versus the dynamical approach in the next section). Adaptations of this description are also (but less so) used by economists, from which ideas that are central to evolution were believed to have been borrowed (including Malthus) by biologists, with some but not all of the key features retained - see Friedman (1998) for more on evolutionary games in economics.

The first arguments in the spirit of evolutionary game theory were given (descriptively) by Charles Darwin in the Descent of Man (1871) in which he details why (from an evolutionary standpoint) the ratio of the sexes should be one-to-one. Shortly after Carl Dúsing gives the first mathematical account of Darwin's sex-ratio argument (see Edwards, 2000), which was continued in Fisher (1930) through Fisher's principle. A solution to Fisher's principle was given by William Hamilton in Hamilton, 1967 (after he was assigned to help Fisher's student Anthony W.F. Edwards), in which he introduces the notion of an "unbeatable strategy". This notion inspired John Maynard Smith and George R. Price, who coined the term Evolutionarily Stable Strategy (ESS) in Maynard Smith and Price (1973). Arguably, ESSs admit the most central topic of evolutionary biology and constitutes the main subject of this report.

In tandem with the idea of an ESS, significant developments in evolutionary games were made in different directions many of which remain active to this day or have laid the foundation for work that is carried out presently. One example includes work that was initiated by Hamilton during the 1960s (see Hamilton, 1964), which was further developed by Robert Trivers in Trivers (1971). The latter emphasised relatedness, altruism and inclusive fitness and successfully explained seemingly curious instances of cooperative behaviour in nature. Another example includes the works of Richard Lewontin, particularly Lewontin (1961) on group selection, which although considered unrealistic can also be thought to have sparked more recent interest in multi-level selection (see Boyd and Richerson, 2002). Finally, Maynard Smith (and less so, Price) besides introducing methods of analysis also invented examples of games that are now considered classic in the realm of evolutionary game theory (such as the Hawk-Dove game and the War of Attrition). Given the breadth of topics that are considered within evolutionary game theory, it makes sense to decide on a starting point

depending on which of these is most relevant to the research question at hand. Our work will be primarily on ESSs and invadability of strategies (although see discussions relating to ecological stability in later chapters), we will naturally take the early works of Maynard Smith and Price as our starting point.

Important work on ESSs was carried out by John Haigh (see Haigh, 1975) shortly after the formal coining of the term by Maynard Smith and Price, in which he introduced a systematic way of identifying all ESSs in a two-player matrix game. Following this, David T. Bishop together with Chris Cannings (still in the 1970s) provided the Bishop-Cannings theorem, which bears important consequences for the co-existence of multiple ESSs (see Bishop and Cannings, 1976). Subsequently, Cannings and Glenn Vickers produced numerous works on the patterns of ESSs (see for example Vickers and Cannings, 1988 and Cannings and Vickers, 1988). An invaluable reference summarising much of this subsequent work on ESSs is Maynard Smith (1982). Up until this point, much of the theory had been applied to linear games (in which the payoff is linear in the strategies - two player matrix games are an example of this), although realistic biological examples (including the model studied in this report) are better described by non-linear games. Notably, theoretical progress on the general theory of non-linear evolutionary games is quite restricted (possibly to Bomze and Pötscher, 1989) although related to non-linear games is the theory of multi-player games which is presently active. This was first developed by Günther Palm in Palm (1984) and has subsequently received attention from Haigh, Cannings and Mark Broom, among others.

Another common simplification in evolutionary game theory is the treatment of populations as being well-mixed and effectively infinite. Realistically, there are circumstances (especially in finite populations) where due to inherent habitat structures, for instance, certain individuals may be more likely to interact with others. Although the evolution of populations with specific structures had been studied beforehand (see for example Maynard Smith and Parker, 1976 or Moran, 1958) it gained popularity when its connection with the existing theory on cellular automata (first studied by Neumann and Stanislaw Ulam in the 1940s) in the Game of Life, which was introduced by Gardner (1970) and developed to consider more general structures using evolution on graphs (see Lieberman et al., 2005 and Nowak, 2006).

1.2 Concepts from classical and evolutionary games

In this section we discuss some of the key principles of game analysis and underpin how these are applied in the biological setting. We will most closely follow the introductory chapters of Broom and Rychtár (2013), although most textbooks on game theory should cover these basics. Readers familiar with these concepts are encouraged to skip this section entirely. We will refer to a game as a mathematical model of a situation in which a collection of entities (finite or infinite) that interact with each other (through conflict or cooperation) by making strategic decisions that influence each other's welfare. A game played by two players is the simplest that may be considered and is commonly used to describe situations in which randomly chosen individuals from a large population of animals engage in pairwise contests (the precise nature of which may vary). It is worth noting at this point that the type of game we consider in later chapters is quite different to the setup described below.

Within a game there may be various points (at least one) at which a player is called upon to make a strategic decision. These decisions constitute what are known as *actions*. A complete specification of the actions to be made at every point in the game is known as a *strategy* and can be viewed as an element drawn from the player's *strategy set*. There is a distinction between *pure* and *mixed strategies* depending on whether players choose any one strategy (from their strategy sets) with certainty. That is, a pure strategy

constitutes a single choice of what strategy to play, whereas a mixed strategy describes the probability with which the players choose any one strategy. More specifically, if a player in a certain game can choose from (finitely many) strategies $(S_1, S_2, ..., S_n)$, one may assign the probability vector $\boldsymbol{p} = (p_1, p_2, ..., p_n)$ to this set so that p_i indicates the probability that strategy S_i is chosen (where $i = 1, 2, ..., n$). It is therefore clear that elements S_i of the strategy set constitute the pure strategies, which we treat as an orthonormal basis spanning our *strategy space*. In particular, any mixed strategy may be expressed as the convex combination (since the coefficients are probabilities that must sum to one)

$$\boldsymbol{p} = (p_1, p_2, ..., p_n) = \sum_{i=1}^{n} p_i S_i. \tag{1.2.1}$$

We proceed by defining the *support* of a mixed strategy.

Definition 1.2.1. (Support) *Let $\boldsymbol{p}$ be a mixed strategy. The* support *of $\boldsymbol{p}$, denoted $S(\boldsymbol{p})$ consists of those indices of pure strategies that have non-zero chance of being played by an individual playing $\boldsymbol{p}$ so that $S(\boldsymbol{p}) := \{i : p_i > 0\}$.*

Notice that this definition applies also to the case for which $\boldsymbol{p}$ is mixed over an infinite set of strategies. An example of this in the War of Attrition, which we detail in the next section of this chapter.

One of the key assumptions of game theory is that the entities make decisions so that the outcomes of those decisions maximise their own welfare. Some of the most famous examples of games are those of the Rock-Scissors-Paper (RSP) game and the Prisoner's Dilemma (PD). In order for the mathematical description of a game to be complete one must specify the rules of the game, including how and when players are to make their decisions (for example in *static* games it is assumed that all players make decisions simultaneously). If the number of pure strategies is infinite (as is the case in the model of Broom et al., 2006 which is introduced in the next chapter) we may identify the strategy set with the interval $[0, \infty)$ and thus use a probability measure $\boldsymbol{p}$ on $[0, \infty)$ such that for any measurable subset $A \subset [0, \infty)$ containing the mixed strategy in question $\boldsymbol{p}(A)$ is the probability that mixed strategy $x \in A$ is picked.

The notion of payoff is central in game analysis and describes the reward received by each player upon choosing a certain strategy. In the context of evolutionary games reward is associated with an individual's level of fitness [2] and the assumption of natural selection necessitates the prevalence of strategies that are associated with higher fitness. It is worth noting that various distinct definitions of fitness have been proposed throughout the history of evolutionary games (see for example Hamilton, 1964 and Dawkins, 1999) showing that a certain choice may be more appropriate than others depending on the assumptions that better suited the situation in question (indeed more specific definitions are introduced in the following chapter). The payoff to a game is usually understood as making up a certain part of the individual's *overall fitness*, the rest being made up by its *background fitness*.

Assume that a game has m players, with $\boldsymbol{S}_i$ being the set of pure strategies available to player i. The fitness of this player is a scalar-valued function f_i, which depends on the strategy choices of all players (including itself). We write

$$f_i : \mathbb{R}^{|\boldsymbol{S}_1|} \times ... \times \mathbb{R}^{|\boldsymbol{S}_m|} \to \mathbb{R} \text{ such that } \boldsymbol{S}_1 \times ... \times \boldsymbol{S}_m \mapsto f_i(\boldsymbol{S}_1, ..., \boldsymbol{S}_m).$$

The *normal form* representation of such a game (as opposed to the *extensive form* representation, which

[2] The expected number of offspring which survive to breeding age is usually a good indicator of fitness - see for instance Broom and Rychtár (2013).

we do not discuss here) involves the complete specification of each player's strategy set together with their payoffs (defined above)

$$\{\boldsymbol{S}_1, ..., \boldsymbol{S}_m; f_1, ..., f_m\}.$$

For the case in which a game involves two players[3] with associated strategy sets $\boldsymbol{S}$ and $\boldsymbol{T}$ the payoff functions

$$f_1 : \mathbb{R}^{|\boldsymbol{S}|} \times \mathbb{R}^{|\boldsymbol{T}|} \to \mathbb{R} \text{ such that } \boldsymbol{S} \times \boldsymbol{T} \mapsto f_1(\boldsymbol{S}, \boldsymbol{T}) \text{ and}$$

$$f_2 : \mathbb{R}^{|\boldsymbol{T}|} \times \mathbb{R}^{|\boldsymbol{S}|} \to \mathbb{R} \text{ such that } \boldsymbol{T} \times \boldsymbol{S} \mapsto f_2(\boldsymbol{T}, \boldsymbol{S}).$$

lead to the more compact normal form representation

$$\{\boldsymbol{S}, \boldsymbol{T}; f_1(\boldsymbol{S}, \boldsymbol{T}), f_2(\boldsymbol{T}, \boldsymbol{S})\}.$$

By introducing indices $i = \{1, ..., |\boldsymbol{S}|\}$ and $j = \{1, ..., |\boldsymbol{T}|\}$ to label the pure strategy components of $\boldsymbol{S}$ and $\boldsymbol{T}$ we can introduce matrix

$$A = (a_{ij})_{i=1,...,|\boldsymbol{S}|, j=1,...,|\boldsymbol{T}|} : \mathbb{R}^{|\boldsymbol{S}|} \times \mathbb{R}^{|\boldsymbol{T}|} \to \mathbb{R},$$

whose (i, j)-component a_{ij} represents the payoff to the first player when he/she plays the i^{th} strategy from its set $\boldsymbol{S}$ and the other plays the j^{th} strategy from its set $\boldsymbol{T}$. Likewise we introduce a second matrix

$$B = (b_{ij})_{i=1,...,|\boldsymbol{T}|, j=1,...,|\boldsymbol{S}|} : \mathbb{R}^{|\boldsymbol{T}|} \times \mathbb{R}^{|\boldsymbol{S}|} \to \mathbb{R} \tag{1.2.2}$$

whose (i, j)-component b_{ij} represents the payoff to the second player when it plays the i^{th} strategy from the set $\boldsymbol{T}$ and the first player plays the j^{th} strategy from its set $\boldsymbol{S}$. It is common practice to convey this payoff information through a so-called *bi-matrix*, whose entries contains the payoff to both players i.e. $(A, B^T)_{i,j} := (a_{ij}, b_{ji})$. That is, the $(i, j)^{th}$ entry of the bi-matrix (A, B^T) gives the payoff to the first and second player when the first chooses the i^{th} from $\boldsymbol{S}$ and the second chooses the j^{th} from $\boldsymbol{T}$.

While matrices A and B completely specify the payoffs to players choosing pure strategies, it remains for us to consider payoffs to players playing mixed strategies. In such cases, there is uncertainty regarding the particular choice of pure strategy the players make and it is thus only sensible for us to consider their *expected payoffs*. If player I plays mixed strategy $\boldsymbol{p}$ and player II plays mixed strategy $\boldsymbol{q}$, the proportion of of games that involve I playing S_i and II playing T_j is $p_i q_j$ and the associated rewards to the players are a_{ij} and b_{ij}. The expected payoffs are achieved by averaging over all such probabilities

$$E_I[\boldsymbol{p}, \boldsymbol{q}] = \sum_{i,j} a_{ij} p_i q_j = \boldsymbol{p} A \boldsymbol{q}^T, \quad \text{and} \tag{1.2.3}$$

$$E_{II}[\boldsymbol{q}, \boldsymbol{p}] = \sum_{i,j} b_{ij} p_i q_j = \boldsymbol{q} B \boldsymbol{p}^T. \tag{1.2.4}$$

Games for which the roles of players I and II are indistinguishable are called *symmetric games*. In such cases it is assumed that players I and II share the same strategy set, say $\boldsymbol{S}$ and furthermore, that the payoff to I when he/she plays strategy S_i and II plays S_j (given by a_{ij}) is the same as the payoff to II when he/she

[3]Such games may be played within much larger populations in which individuals engage in pairwise contests

plays S_i and I plays S_j (given by b_{ji}). That is, in this situation we have $a_{ij} = b_{ji}$ for all $i, j \in \{1, ..., |\boldsymbol{S}|\}$, which holds if and only if $\mathrm{B} = \mathrm{A}^T$. Simple algebra can show that for symmetric games $E_I[\boldsymbol{p}, \boldsymbol{q}] = E_{II}[\boldsymbol{p}, \boldsymbol{q}]$, it therefore only makes sense to consider payoffs to player I (as that of player II can be directly obtained), which we denote by $E[\boldsymbol{p}, \boldsymbol{q}]$ without subscript.

It is common in evolutionary games to consider large (effectively infinite) populations (denoted Π) made up of individuals playing different strategies; a natural way of describing their structure is to consider the density of individuals playing any one strategy. A *homogeneous* population is one that consists of (almost all) individuals playing a certain strategy $\boldsymbol{p}$ and which can be described using the delta distribution $\Pi = \delta_{\boldsymbol{p}}$. In these, the probability that a randomly selected individual plays $\boldsymbol{p}$ is precisely one. If there is a finite number of pure strategies to be chosen from, δ_i (naturally) denotes a population consisting of individuals all playing strategy S_i and thus in an *inhomogeneous* population we let p_i denote the proportion of individuals playing S_i. Recall that in (1.2.1) we allowed mixed strategies to be represented as convex combinations of pure strategies S_i, which we subsequently identified as an orthonormal basis (of row vectors) spanning the strategy simplex. Similarly, we identify population structure δ_i with the column vector S_i^T and express composite population structures as convex combinations of these, so that

$$\Pi = \sum_i p_i \delta_i \tag{1.2.5}$$

describes a population in which the proportion of individuals playing S_i is p_i.

The analogy with mixed strategies may be extended further so that the column vector $\boldsymbol{p}^T = \sum_i p_i \delta_i$ denotes a mixed population, while the row vector $\boldsymbol{p} = \sum_i p_i S_i$ denotes a mixed strategy. An interesting connection between the two analogies is that $\delta_{\boldsymbol{p}}$ describes a uniform population playing mixed strategy $\boldsymbol{p}$, while $\boldsymbol{p}^T = \sum_i p_i \delta_i$ denotes a mixed population playing pure strategies p_i. In this sense $\delta_{\boldsymbol{p}}$ describes the average strategy of a mixed population and is analogous to the barycenter of a rigid body in mechanics. A *focal* individual in the mixed population would play (a matrix game) against randomly chosen individual S_i with probability p_i, which is the same as if the individual were in the uniform population $\delta_{\boldsymbol{p}}$.

Of central importance is the function $\mathcal{E}[\sigma; \Pi]$ which represents the fitness of an individual playing strategy σ in a population represented with Π. As with mixed strategies we are interested in the expected payoff (the exact reward is hard to evaluate) to an individual playing (pure or mixed) strategy σ in a population $\Pi = \sum_i p_i \delta_i$. For simplicity, we consider matrix games. Assume that the focal individual plays k games with randomly-chosen opponents from the population, whose strategies cannot be known in advance. The expected number of games against an individual playing S_i is $p_i k$ and the total reward from these encounters is

$$\sum_i p_i k E[\sigma, S_i] = \sum_i p_i k \sum_j \sigma_j a_{ji},$$

and the average payoff from one such encounter is obtained by dividing through by k, which gives

$$\mathcal{E}[\boldsymbol{\sigma}, \boldsymbol{p}^T] = \sum_i p_i \sum_j \sigma_j a_{ji} = \boldsymbol{\sigma} A \boldsymbol{p}^T = E[\boldsymbol{\sigma}, \boldsymbol{p}]. \tag{1.2.6}$$

Given that the quantity $E[\sigma, \boldsymbol{p}]$ is bilinear, it follows that $\mathcal{E}[\sigma, \boldsymbol{p}^T]$ is linear both in the strategies of the focal individual and also in the composition of the population $\boldsymbol{p}^T$. This is an important consequence and one that allows important results to be established both in the static and in the dynamic approaches, which we briefly describe below.

Two approaches in evolutionary game theory

When modelling change in populations it is important to notice that not all processes evolve on the same time-scale. It is therefore natural to adopt different notions of stability depending on the types of dynamics observed (see Hofbauer and Sigmund, 1998 for a systematic review). There exists a (generally accepted) hierarchy of time-scales governing the various processes, which are usually taken to evolve independently. If a fast-paced and a slower process evolve simultaneously, the latter is taken to be fixed when focus is on the former, while when the slower process is considered the fast one is assumed to be in equilibrium. In particular, it is assumed that animal behaviours (including resting and foraging strategies) change rapidly, while *population dynamics* (describing how the frequencies of individuals playing a certain strategy change over successive generations as a result of natural selection) is slower than the behavioural dynamics, although still considered short-term. Furthermore, the *evolutionary dynamics* describing how new mutant strategies can invade an existing population is even slower (long-term). Finally, beneficial mutations are so rare that they are assumed to occur only after competition with a previous mutant has been concluded. Mutant strategies are therefore (more questionably) introduced 'one-at-a-time' and in this sense, mutation is an even slower process than the evolutionary dynamics.

Interestingly, the relevant literature in evolutionary games can be broadly categorised into those works that follow a dynamical approach versus those that follow a static one. The main difference between these is the time-scale of the processes considered, with dynamical approaches focusing more on the short-term (including animal behaviour and population dynamics) and static approaches focusing on the long-term (including evolutionary dynamics and mutation) without paying great attention to how replacement of one strategy by another takes place. In the second part of the report (chapters 5 and 6) we consider stability on two different time-scales and determine the conditions required to achieve stability in a more general sense.

The dynamical approach has seen much development over the past decades; it has subsequently been subdivided into various strands with the *replicator dynamics* maintaining focal importance. The replicator dynamics considers a population $\Pi = \sum_i p_i \delta_i = \boldsymbol{p}^T$ of size N in which the number of individuals playing S_i is $N_i = p_i N$ and are assigned fitness $f_i(\boldsymbol{p})$. The key assumption here is that the size of a group playing a certain strategy grows in proportion to its fitness. In the discrete case in which generations do not overlap and reproduction is asexual (a reasonable assumption for studying a number of insect species and/or annual plants), we have

$$N_i(t+1) = N_i(t) f_i(\boldsymbol{p}(t)),$$

from which it follows that

$$p_i(t+1) = p_i(t) \frac{f_i(\boldsymbol{p}(t))}{\bar{f}_i(\boldsymbol{p}(t))}, \tag{1.2.7}$$

where $\bar{f}(\boldsymbol{p}) = \sum_i p_i f_i(\boldsymbol{p})$ is the average fitness of the population[4]. In the continuous replicator dynamics (with overlapping generations and asexual reproduction) a similar assumption which is key holds, namely

$$\dot{N}_i(t) = N_i f_i(\boldsymbol{p}(t))$$

from which it can be shown that

$$\dot{p}_i(t) = p_i(t)\big(f_i(\boldsymbol{p}(t) - \bar{f}(\boldsymbol{p}(t))\big). \tag{1.2.8}$$

Comparing (1.2.7) and (1.2.8) it is clear that the density of individuals playing S_i grows in proportion to the

[4]Recall that in the case of pairwise contests the fitness can be described in terms of a payoff matrix

fitness of that strategy compared with the average fitness of the population. Further, these equalities make it clear that fixed points and further stability characterisations can be employed very much in the spirit of dynamical systems theory.

The replicator dynamics study the replacement of one strategy by another by looking at the variations in the frequency of individuals in the population who play a a given strategy. An important distinction to remark here is that the set of potential strategies from which prey can choose is fixed and pre-determined[5]. The latter constitutes the object of study of *adaptive dynamics*, which typically considers population compositions of the form $\Pi = (1-\varepsilon)\boldsymbol{p} + \varepsilon \boldsymbol{p}'$, involving a majority playing the *resident strategy* $\boldsymbol{p}$ and a *mutant* minority playing a strategy p' that is local to $\boldsymbol{p}$ such that $\boldsymbol{p} \approx \boldsymbol{p}'$ and $0 < \varepsilon \ll 1$. If the fitness of individuals playing the mutant strategy is greater than the fitness of individuals playing the resident strategy then we say that the mutants have the potential to *invade* the residents. The time-scale of the described dynamics is slower compared with the replicator dynamics and the concepts of stability employed are closer in spirit to those employed later in the report. Indeed, a key concept in adaptive dynamics is that of an *Evolutionarily Singular Strategy*, which describes a strategy for which no local mutation has higher associated fitness. A central component of this report is the study of static properties associated with the evolutionary model proposed by Broom et al. (2006) in hope that neglecting the underlying dynamics will not have a significant impact on our conclusions.

1.3 Static concepts in evolutionary game theory

In a static approach the specific faster-evolving dynamics (including changes in animal behaviour and/or population dynamics) are considered constant so that the quality of a strategy (*best* or *worst*) is determined solely through a game-theoretical standpoint. In practice, this is achieved by first assuming that the population can be described in terms of a certain strategic composition and by hence determining whether there is incentive for this composition to change (i.e. for players to switch to strategies with higher associated fitness). Therefore, how a certain population composition can be achieved is not an immediate concern when adopting a static approach. From the purely game-theoretical standpoint, equilibrium points are identified by assuming intelligence among players: the best strategy for one player is the best response strategy to the other player's best strategy, where the other player's best strategy is his/her best response to the first player's best strategy. In particular, we have the following definition (due to Nash)

Definition 1.3.1. (Best response) *A strategy S is a best response to strategy T if*

$$f(S', T) \leq f(S, T) \text{ for all strategies } S', \tag{1.3.1}$$

where f(S,T) denotes the payoff to player using S against a player using T.

This leads us to the definition of a Nash Equilibrium (N.E.), which is describes a point in the strategy space at which players have no incentive to change their strategies

[5] An evident restriction of the adaptive dynamics description is that it does not account for the emergence of new strategies

Definition 1.3.2. (Mixed Nash equilibrium) *In a two-player game in which players are allowed to play mixed strategies, the pair* $(\boldsymbol{p}^*, \boldsymbol{q}^*)$ *is a (mixed) Nash equilibrium if* $\boldsymbol{p}^*$ *is best response to* $\boldsymbol{q}^*$ *and vice-versa. That is, if*

$$E_I[\boldsymbol{p}^*, \boldsymbol{q}^*] \geq E_I[\boldsymbol{p}, \boldsymbol{q}^*] \quad \textit{for all} \quad \boldsymbol{p} \neq \boldsymbol{q}^*, \ \textit{and}$$
$$E_{II}[\boldsymbol{q}^*, \boldsymbol{p}^*] \geq E_{II}[\boldsymbol{q}, \boldsymbol{p}^*] \quad \textit{for all} \quad \boldsymbol{q} \neq \boldsymbol{q}^*, \tag{1.3.2}$$

where $E_i[\boldsymbol{p}, \boldsymbol{q}]$ *denotes the payoff to player* i *if he/she plays strategy* $\boldsymbol{p}$ *while the other plays* $\boldsymbol{q}$*.*

Notice that for symmetric games this definition amounts to any player playing the best-response against him/her self, so that $(\boldsymbol{p}^*, \boldsymbol{p}^*)$ is a Nash equilibrium if

$$E[\boldsymbol{p}, \boldsymbol{p}^*] \leq E[\boldsymbol{p}^*, \boldsymbol{p}^*] \quad \textit{for all} \quad \boldsymbol{p} \neq \boldsymbol{p}^*. \tag{1.3.3}$$

In the framework of biology one does not explicitly assume intelligence among players but tends to assume that strategies/traits evolve in the context of natural selection. In particular, if an individual plays strategy S as a result of evolution then any other strategy played by the remaining players will also be a result of evolution. It is therefore clear that the best strategy must again be a best response strategy against itself. In fact, if everybody in the population were to play S, so that $\Pi = \delta_S$ and this were not the best response to itself then a single individual playing some different strategy $M \neq S$ would have higher payoff than when playing S. Nonetheless, imposing that a strategy is best response to itself does not necessarily imply that if the population were to play it that it would be stable. Rather, for a population to be stable while playing $\boldsymbol{S}$ that they have higher payoff than that of a mutant group playing some competing strategy M. In particular, we consider composition $\Pi = (1-\varepsilon)\delta_S + \varepsilon\delta_M$ and hence arrive at the following definitions

Definition 1.3.3. (Evolutionarily Stable Strategy) *We say that a strategy* S *is evolutionarily stable against strategy* M *if there exists* $\varepsilon_M > 0$ *so that for all* $\varepsilon < \varepsilon_M$ *we have*

$$\mathcal{E}[S; (1-\varepsilon)\delta_S + \varepsilon\delta_M] > \mathcal{E}[M; (1-\varepsilon)\delta_S + \varepsilon\delta_M]. \tag{1.3.4}$$

Furthermore, S *is an* Evolutionarily Stable Strategy (ESS) *if it is evolutionarily stable against* M *for every other strategy* $M \neq S$*.*

An ESS is a strategy that cannot be invaded by any other strategy if the associated mutant group was below a certain size. In general, this threshold is different for every strategy, and we thus arrive at yet a stronger condition than that of an ESS, which is that of a strategy being (uniformly) uninvadable.

Definition 1.3.4. (Uniform Invasion Barrier) *A strategy* $\boldsymbol{p}$ *is* uniformly uninvadable *if there exists* $\varepsilon_{\boldsymbol{p}} > 0$ *such that for every other strategy* $\boldsymbol{q} \neq \boldsymbol{p}$ *and for all* $0 < \varepsilon < \varepsilon_{\boldsymbol{q}}$ *we have*

$$\mathcal{E}[\boldsymbol{p}; (1-\varepsilon)\delta_{\boldsymbol{p}} + \varepsilon\delta_{\boldsymbol{q}}] > \mathcal{E}[\boldsymbol{q}; (1-\varepsilon)\delta_{\boldsymbol{p}} + \varepsilon\delta_{\boldsymbol{q}}]. \tag{1.3.5}$$

Stability theory is well-developed for matrix games and we detail the adaptations of the general theory here. From (1.2.6) we have that $\mathcal{E}[\sigma, \boldsymbol{p}^T] = E[\sigma, \boldsymbol{p}]$ and we therefore have the following natural adaptation of the above definition.

Definition 1.3.5. (ESS in matrix games) *A (pure or mixed) strategy $\boldsymbol{p}$ is an ESS for a matrix game if for every other strategy $\boldsymbol{q} \neq \boldsymbol{p}$ there exists $\varepsilon_{\boldsymbol{q}} > 0$ such that for all $0 < \varepsilon < \varepsilon_{\boldsymbol{q}}$, we have*

$$E[\boldsymbol{p}, (1-\varepsilon)\boldsymbol{p} + \varepsilon\boldsymbol{q}] > E[\boldsymbol{q}, (1-\varepsilon)\boldsymbol{p} + \varepsilon\boldsymbol{q}]. \tag{1.3.6}$$

This definition was a direct adaptation to the two-player setup of the general definition provided earlier. A more useful definition is the following, which can be shown to be equivalent to the above (see for example Broom and Rychtár, 2013).

Definition 1.3.6. (ESS in matrix games #2) *A (pure or mixed) strategy $\boldsymbol{p}$ is an ESS for a matrix game if and only if for any mixed strategy $\boldsymbol{q} \neq \boldsymbol{p}$ we have*

$$E[\boldsymbol{p}, \boldsymbol{p}] \geq E[\boldsymbol{q}, \boldsymbol{p}] \text{ and if } E[\boldsymbol{p}, \boldsymbol{p}] = E[\boldsymbol{q}, \boldsymbol{p}], \text{ then } E[\boldsymbol{p}, \boldsymbol{q}] > E[\boldsymbol{q}, \boldsymbol{q}]. \tag{1.3.7}$$

For matrix games in particular, the connection between and ESS and a uniform invasion barrier is straightforward, as indicated by the following theorem (for a proof see Broom and Rychtár, 2013)

Theorem 1.3.7. *For a matrix game, strategy $\boldsymbol{p}$ is an ESS if and only if it is uniformly uninvadable, namely there exists $\varepsilon_{\boldsymbol{p}} > 0$ such that for all $\boldsymbol{q} \neq \boldsymbol{p}$ and for all $0 < \varepsilon < \varepsilon_{\boldsymbol{p}}$ we have*

$$\mathcal{E}[\boldsymbol{p}; (1-\varepsilon)\delta_{\boldsymbol{p}} + \varepsilon\delta_{\boldsymbol{q}}] > \mathcal{E}[\boldsymbol{q}; (1-\varepsilon)\delta_{\boldsymbol{p}} + \varepsilon\delta_{\boldsymbol{q}}]. \tag{1.3.8}$$

There is another important result for mutations that are local, namely (once more for a proof of this result the reader is encouraged to consult Broom and Rychtár, 2013)

Theorem 1.3.8. (Local superiority) *For a matrix game, $\boldsymbol{p}$ is an ESS if and only if*

$$E[\boldsymbol{p}, \boldsymbol{q}] > E[\boldsymbol{q}, \boldsymbol{q}] \text{ for all } \boldsymbol{q} \neq \boldsymbol{p} \text{ sufficiently close to } \boldsymbol{p}. \tag{1.3.9}$$

Finally, recalling the definition of the support $S(\boldsymbol{p})$ of a mixed strategy as given in definition 1.2.1, we present the following lemma, which we apply in the examples that follow. This says that the pure strategies in the support of $\boldsymbol{p}$ do equally well against $\boldsymbol{p}$ as does $\boldsymbol{p}$ against itself. In particular, we have the following.

Lemma 1.3.9. *Let $\boldsymbol{p} = (p_i)$ be an ESS. Then for any $i \in S(\boldsymbol{p})$ we have $E[S_i, \boldsymbol{p}] = E[\boldsymbol{p}, \boldsymbol{p}]$.*

Motivated by this, we may identify those indices of an arbitrary mixed strategy $\boldsymbol{q}$ associated with the pure strategies that do equally well against $\boldsymbol{p}$ as does $\boldsymbol{p}$ against itself and so for this, we write $T(\boldsymbol{q}) := \{i : E[S_i, \boldsymbol{p}] = E[\boldsymbol{p}, \boldsymbol{p}]\}$. For arbitrary mixed strategy $S(\boldsymbol{p}) \subseteq T(\boldsymbol{p})$, while if $\boldsymbol{p}$ is an ESS then $S(\boldsymbol{p}) = T(\boldsymbol{p})$ as suggested by lemma 1.3.9. Furthermore, we have the following lemma (for a proof of which see Broom and Rychtár, 2013).

Lemma 1.3.10. *Let $\boldsymbol{p}$ be an ESS. Then $E[\boldsymbol{q}, \boldsymbol{p}] = E[\boldsymbol{p}, \boldsymbol{p}]$ if and only if $S(\boldsymbol{q}) \subseteq T(\boldsymbol{p})$.*

We close this section by presenting two further consequences about ESSs (with proofs once again given in Broom and Rychtár, 2013), both of which are especially useful for ruling out ESSs.

Theorem 1.3.11. (Bishop and Cannings 1976) *If $\boldsymbol{p}$ is an ESS of a matrix game and there exists mixed strategy $\boldsymbol{q} \neq \boldsymbol{p}$ with $S(\boldsymbol{q}) \subseteq T(\boldsymbol{p})$ then $\boldsymbol{q}$ is not an ESS of that game.*

This final result follows as a corollary of the above theorem of Bishop and Cannings and will prove useful in the analysis of the War of Attrition in the analysis that follows.

Remark 1.3.12. *Let $\boldsymbol{p}$ be an internal ESS (with $p_i > 0$ for all i) of a matrix game. Then $\boldsymbol{p}$ is unique.*

Classic Examples of Games

We devote this part to the discussion of two classical examples of evolutionary games, namely the Hawk-Dove game and the War of Attrition. Of course, there are many, many others such as the Prisoner's Dilemma, the Rock-Scissors-Paper or the Sex Ratio game, but unfortunately their presentation would exceed our scope. We have chosen the Hawk-Dove game, merely to demonstrate the general ideas discussed thus far in the context of a straightforward, two-player matrix game and the War of Attrition as an example involving continuous traits which is an example of *playing the field*. The central game of this thesis is non-linear (differences in the value of the trait are not proportional to the associated differences in the payoff) and yet despite this exhibits a number of similarities (ESSs are manifest as continua in the strategy space) with the War of Attrition, which is a typical example of a game that is linear.

The Hawk-Dove game

The Hawk-Dove game was introduced in Maynard Smith and Price (1973) and has seen numerous variations since - here we look at its basic formulation. It is was originally known for describing animal contests over territory (such as that played by stags during the breeding season - see Broom and Rychtár (2013)) although it may also model more general situations in which contestants are called upon to decide whether to fight (play *Hawk*) or to flee (play *Dove*). Assuming a population made up of identical individuals (in terms of strength and related attributes), the game is taken to be symmetric with the associated payoffs provided by the bi-matrix

	Hawk	Dove
Hawk	$\left(\frac{V-C}{2}, \frac{V-C}{2}\right)$	$(V, 0)$
Dove	$(0, V)$	$\left(\frac{V}{2}, \frac{V}{2}\right)$.

In particular, if a Hawk encounters a Dove, the former is victorious (V) and gains the full reward (i.e. control over territory) and the latter gains nothing and backs away without fighting. If a Hawk encounters a Hawk, they both share the same reward and both bear a certain cost ($C/2$) on account of engaging in conflict. Finally, if a Dove encounters a Dove then they both share the reward (control over territory) equally ($V/2$) and without baring any costs (as they do not fight). The static analysis identifies three cases, depending on whether the reward outweighs the cost of conflict over this namely $V > C$, $V = C$ (this is non-generic) and $V < C$. For the first, (H, H) is the unique (pure) N.E.. Furthermore, it is strict and symmetric and thus the associated strategy Hawk is an ESS. In the second case ($V = C$), we can identify three (pure) N.E., namely $(H, H), (H, D)$ and (D, H) of which only the former is symmetric. In particular, $E(H, H) = E(D, H)$ but since $E(H, D) > E(D, D)$ then by Definition (1.3.6), Hawk is still an ESS. In the latter case ($V < C$), the two (pure) N.E. are (D, H) and (H, D) are non-symmetric, from which we deduce that there is no pure ESS. It can be shown that mixed strategy $\boldsymbol{p} = (V/C, 1 - V/C)$ is the unique ESS when $V < C$. In particular, this shows that the H-D game has a unique ESS, irrespective of the values of V and C.

For the sake of comparison with the static it is worth mentioning the applicability of the dynamical approach to a population consisting of individuals engaging in pair-wise Hawk-Dove contests. For the

continuous replicator dynamics we substitute the fitness in (1.2.8) with the expected payoff (in terms of the payoff matrix given above) and by set $\boldsymbol{p} = (p, 1-p)$, with p representing the proportion of individuals playing Hawk. We may visualise the dynamics through a phase portrait and thus characterise ESSs as attractors of the dynamics - a very different approach to the static one, yet yielding the same conclusions about optimal strategy.

The War of Attrition

The War of Attrition is an example of *playing the field*. As before, this is a game in which two individuals are competing over a reward, but now the players can choose from an infinite (and uncountable) number of strategies. The players each display themselves for some time, the first to leave gets no reward and the one who waits the longest gets the full reward. In this situation there is no physical cost incurred from playing the game, but there is a certain opportunity cost associated with waiting, on account of the fact that the time spent waiting could have been utilised (more productively) otherwise. Another assumption is that the game is static, namely that the players decide their strategies before engaging and cannot change them subsequently.

A pure strategy is denoted by S_t and is played by an individual who is prepared to wait for a time of exactly $t \geq 0$. A mixed strategy $\boldsymbol{p}$ is a probability measure on the set $[0, \infty)$, and is much like a density function $p(x)$ defined on $x \geq 0$ in the sense that the probability of an individual leaving between times x and $x + dx$ is $p(x)dx$ and the probability of leaving before time t is $\int_0^t p(x)dx$. We mention for completeness that a pure strategy S_t can be seen as a Dirac measure at the point t. Indeed, any continuous function $f : [0, +\infty) \to \mathbb{R}$ can be expressed as

$$f(t) = \int_0^\infty f(x) dS_t(x).$$

The payoff is the reward minus the cost of waiting for the reward, where the former is assumed to be a fixed positive constant V and the cost $C(x)$ is assumed to be proportional to the time spent waiting so that $C(x) := cx$ for some fixed constant c. In particular, the payoff to individuals playing pure strategies S_x and S_y is given by

$$E[S_x, S_y] = \begin{cases} V - cy, & x > y \\ V/2 - cx, & x = y \\ -cx & x < y, \end{cases} \tag{1.3.10}$$

whereas for mixed strategies $\boldsymbol{p}$ and $\boldsymbol{q}$ (and associated density functions $p(x)$ and $q(x)$) we have

$$E[\boldsymbol{p}, \boldsymbol{q}] = \int\int_{(x,y)\in[0,\infty)^2} E[S_x, S_y] d\boldsymbol{p}(x) d\boldsymbol{q}(y). \tag{1.3.11}$$

From the above payoff functions and from Definition 1.3.6 it is concluded that there are no pure ESSs. Consider the payoff to a player playing pure strategy S_t for some $t > 0$. Then, for any $\tau > 0$, from (1.3.10) we deduce that

$$E[S_{t+\tau}, S_t] = V - ct > \frac{V}{2} - ct = E[S_t, S_t], \tag{1.3.12}$$

and by the equivalence of Definition 1.3.6 and Definition 1.3.5 this is equivalent to saying that a minority of

$S_{t+\tau}$ players can invade a majority of S_t players. Namely that for some small $\varepsilon > 0$

$$\mathcal{E}[S_{t+\tau}; (1-\varepsilon)\delta_{S_t} + \varepsilon\delta_{S_{t+\tau}}] > \mathcal{E}[S_t; (1-\varepsilon)\delta_{S_t} + \varepsilon\delta_{S_{t+\tau}}]. \tag{1.3.13}$$

So as to find suitable ESS candidates, we must therefore restrict our attention to mixed strategies. The support of a mixed strategy was introduced in Definition 1.2.1 and can be extended for this case, in which players may choose from an infinite (and uncountable) number of strategies. Namely, the support of mixed strategy $\boldsymbol{p}$ with associated density function $p(t)$ is precisely the set $S(\boldsymbol{p}) = \{t \geq 0 : p(t) > 0\}$. Adapting lemma 1.3.9 appropriately, we conclude that if $\boldsymbol{p}$ is an ESS then it must satisfy

$$E[S_t, \boldsymbol{p}] = E[\boldsymbol{p}, \boldsymbol{p}] \quad \text{for almost all } \ t \in S(\boldsymbol{p}). \tag{1.3.14}$$

The LHS of this equality can be evaluated using definition 1.3.11

$$\begin{aligned} E[S_t, \boldsymbol{p}] &= \int\int_{(x,y)\in[0,\infty)^2} E[S_y, \boldsymbol{p}] dS_t(y) dp(x) = \int_{x\in[0,\infty)} p(x)dx \int_{y\in[0,\infty)} E[S_y, \boldsymbol{p}] S_t(y) \\ &= \int_{x\in[0,\infty)} E[S_t, \boldsymbol{p}] p(x)dx = \int_0^t E[S_t, \boldsymbol{p}] p(x)dx + \int_t^\infty E[S_t, \boldsymbol{p}] p(x)dx \end{aligned}$$

In light of (1.3.10) and the above, equality (1.3.14) reads

$$\int_0^t (V - cx)p(x)dx + \int_t^\infty (-cxt)p(x)dx = E[\boldsymbol{p}, \boldsymbol{p}]. \tag{1.3.15}$$

Differentiating both sides with respect to t (and noting that the RHS does not depend on t) we deduce that

$$(V - ct)p(t) - c\int_t^\infty p(x)dx + ctp(t) = 0 \Leftrightarrow Vp(t) - c\int_t^\infty p(x)dx = 0. \tag{1.3.16}$$

Setting $p(x) = P'(x)$ on all $x \geq 0$ for some function P implies that the above amounts to

$$VP'(t) - c\int_t^\infty P'(x)dx = 0 \Leftrightarrow P'(t) = \frac{c}{V}P(t) \Leftrightarrow P(t) = K\exp\left(-\frac{c}{V}t\right),$$

for some constant $K \in \mathbb{R}$. From $p(t) = P'(t)$, we deduce that

$$p(t) = -K\frac{c}{V}\exp\left(-\frac{c}{V}t\right) \tag{1.3.17}$$

solves (1.3.15) for almost all t with $p(t) \neq 0$ and imposing that the constant K is chosen so that $\boldsymbol{p}$ is a probability measure, i.e. so that $\int_0^\infty p(x)dx = 1$ gives $K = 1$. It should be remarked that while we have demonstrated that (1.3.17) is an ESS, we have not shown it is unique. For instance, one could conclude that

$$p_{ESS}(t) = \frac{c}{V}\exp\left(-\frac{c}{V}t\right) \quad \text{and} \quad p(t) = \begin{cases} 0, & t \in [0,1) \\ \dfrac{c}{V}\exp\left(-\dfrac{c}{V}(t-1)\right), & t > 1 \end{cases} \tag{1.3.18}$$

are both ESSs since they satisfy (1.3.15) and the necessary condition for normalisation $\int_0^\infty p(t)dt = 1$. To demonstrate that (1.3.17) is the unique ESS we evoke a result from Bishop and Cannings (1976). The theorem is as follows. Let $\boldsymbol{q}$ and $\boldsymbol{r}$ be two different mixed strategies with associated densities admitting differences on sets of non-zero measure (i.e. their difference is non-trivial). The result due to Bishop and Cannings (1976) states that for these we must have

$$E[\boldsymbol{r}, \boldsymbol{r}] - E[\boldsymbol{q}, \boldsymbol{r}] - E[\boldsymbol{r}, \boldsymbol{q}] + E[\boldsymbol{r}, \boldsymbol{r}] < 0 \quad \text{for all} \quad \boldsymbol{q} \neq \boldsymbol{r}. \tag{1.3.19}$$

Notice that at most one ESS can satisfy the above inequality, the reason being that if $\boldsymbol{q}$ cannot invade $\boldsymbol{r}$ (namely $E[\boldsymbol{r}, \boldsymbol{r}] > E[\boldsymbol{q}, \boldsymbol{r}]$) then $\boldsymbol{r}$ invades $\boldsymbol{q}$ ($E[\boldsymbol{r}, \boldsymbol{q}] > E[\boldsymbol{q}, \boldsymbol{q}]$). Further, from (1.3.14) it follows that $\boldsymbol{P}_{ESS}$ must satisfy

$$E[\boldsymbol{p}_{ESS}, \boldsymbol{p}_{ESS}] = E[\boldsymbol{q}, \boldsymbol{p}_{ESS}] \tag{1.3.20}$$

for any pure strategy $\boldsymbol{q} = S_t$ with $t \geq 0$ since $\boldsymbol{p}_{ESS}$ has positive density $p_{ESS}(t) > 0$ for all $t \geq 0$. Identifying $\boldsymbol{r}$ with $\boldsymbol{p}_{ESS}$ and $\boldsymbol{q}$ as described, inequality 1.3.19 now implies

$$E[\boldsymbol{p}_{ESS}, \boldsymbol{q}] > E[\boldsymbol{q}, \boldsymbol{q}] \quad \text{for all} \quad \boldsymbol{q} \neq \boldsymbol{p}_{ESS}, \tag{1.3.21}$$

which identifies $\boldsymbol{p}_{ESS}$ as the unique ESS of the game.

Chapter 2

Aposematic Defence

In this chapter we discuss how prey individuals defend themselves against attack from potential predators and focus on those changes in their visual appearance that achieve this. Depending on the time at which defence mechanisms are activated relative to encounter, these serve a different purpose and may bear different costs to an animal that acquires them. We focus on a particular instance of defence in which prey individuals invest in unpalatable toxins that are transmitted to predators during or after encounter but whose presence is signalled beforehand through bright colourations. This phenomenon is called *aposematism* and is curious from an evolutionary standpoint as it seems to provide no direct benefit to the prey individuals that acquire it upon first inspection. The chapter opens with some background into prey defences and aposematism, including a brief overview of some existing mathematical approaches in aposematism. In the second section, we introduce the game-theoretical description of Broom et al. (2006). We distinguish between the more general version of this model and the prey-predator description, which will hence occupy the remaining portion of the thesis. We present and prove a number of fascinating results from Broom et al. (2006), which relate to the evolutionary stability of aposematism and form the foundation for later chapters.

2.1 Background

The purpose of this section to provide some brief background into the biology and mathematical modelling of aposematism, with emphasis in those areas that are more relevant to the game-theoretical description of Broom et al. (2006).

The biology of aposematism

Organisms from anywhere within the tree of life exhibit defence mechanisms that are deployed to prevent potential predators from mounting attacks. These exist in a variety of guises, are manifested in a range of different ways, trigger the senses differently and pose different fitness advantages to those who deploy them (see Ruxton et al., 2019 for a more elaborate overview of this topic). Defences can be effectively classified depending on whether they are permanently present (*static constitutive*) in the prey individuals that acquire them or whether these are deployed during conflict (*induced*). While the latter have lower costs associated with maintenance and may be better suited against attacks that take place over a longer time period, they tend to be less effective against the type of attack that we consider here (mounted at a fast pace), particularly because the deployment of induced defences typically takes time. The reader is directed to Irie and Iwasa

(2005) or Shudo and Iwasa (2001) for model-based studies, to the case study by Hammill et al. (2008) analysing the costs of induced defences on water fleas (*Daphnia pulex*) or finally to the more recent study by Boots and Best (2018) on the evolution of induced and constitutive defences against infectious diseases. The focus of this report is on rapid (and potentially lethal) attacks and considers mechanisms of defence that are most relevant in this context.

On a more basic level, prey defend themselves against attacks by reducing the chance of that they are detected or encountered by means of *primary defences*, which include *crypsis* (resemblance to the background) or *masquerade* (resemblance to items that are of no inherent interest to the predator such as leaves, twigs or stones). If an attack has been mounted by a predator, prey may still be able to defend themselves through *secondary defences*, whose role is typically two-fold: to reduce the probability that a mounted attack is lethal and to reduce the probability that similar-looking prey are attacked in the future (see Gols, 2014 for a breakdown of *"direct"* and *"indirect"* defences in a large array of chemically-defended plants or chapter 8 of Dickinson et al. (1981) titled *apetite-aversive interactions and inhibitory processes* by Dickinson and Boakes).

Secondary defences are present in a vast range of taxonomic groups throughout the tree of life and are remarkably variable; we coarsely classify these as follows. Prey may possess *chemical* defences (such as olfactory deterrents, repellent secretions or internally stored toxins that are transferred to the predator either after or during encounter), *mechanical* defences (which could include the possession of sharp spines, tough integuments, the ability to change shape) or *behavioural* defences (including aggressive retaliation or social-defensive grooming). We should remark that while some forms are readily visible by predators at a distance, other forms are not and in those cases the defence is often paired with additional signalling cues. On p.232 of Wallace (1889) it is remarked for such prey that they "*require some signal or danger flag which shall serve as a warning to would-be enemies not to attack them, and they have usually obtained this in the form of conspicuous or brilliant coloration, very distinct from the protective tints of the defenceless animals allied to them*". This is often observed in chemically-defended prey including many species from the Poison dart (*Dendrobatidae*) frog family (see Summers and Clough, 2001a for a discussion on the evolution of colouration and toxicity or the more recent study by Barnett et al., 2018, emphasising how protective colouration triggers a variety of different visual systems at various distances).

Aposematism is the signalling by prey individuals (animals or plants) to predators that they are unprofitable to consume. A succinct description of the process is provided in Wallace (1877), p. 651: *"...warning colours are exceedingly interesting, because the object and effect of these is, not to conceal the object, but to make it conspicuous. To these creatures it is useful to be seen and recognised, the reason being that they have a means of defence which, if known, will prevent their enemies from attacking them, though it is generally not sufficient to save their lives if they are actually attacked."* Indeed, these warning colourations or signals are associated with some form of prey defence and may be manifest in a wide range of physical characteristics perceivable to the predator through sensory stimuli beyond just sight but which may also include smell, touch, taste or combinations of these (see Rowe and Guilford, 1999 for a discussion on the evolution of *"multimodal warning displays"* or Yack (2022) discussing the more poorly understood acoustic defence mechanisms deployed by caterpillars). The term *aposematism* literally means *to keep predators at a distance* and stems from the Greek *apostasis* (meaning *distance*) and *sima* (*signal*) and was coined in Poulton (1890) as an adaptation that more explicitly describes a warning colouration as a signal that warns predators off[1].

From this description we establish that the process of warning colouration may effectively be phrased

[1]For a very brief, up-to-date description of aposematism a good first option is Rojas et al. (2021)

as "aposematic prey sacrifice their primary defence in favour of a signalling appearance that (i) signals to predators the presence of a secondary defences and which in turn (ii) acts as a deterrent and hence a substituted form of primary defence." Indeed, aposematism is a complex process involving the coordinated (and often synchronous) mobilisation of primary and secondary mechanisms. It may come as no surprise that in most taxonomic groups aposematic individuals are rare compared with their camouflaged counterparts - see Santos et al. (2003), Vences et al. (2003) and Ruxton et al., (2019). Chemical defences (in the form of toxins stored within the body) are typically not detectable by the predator at a distance; their presence is perceived only after an attack has been mounted, which distinguishes them from a large group of defences (although behavioural defences including ability flee or fight back also fall under this category).

It is common to refer to an aposematic signal as a phenotype that informs potential predators of the presence of a defence that may otherwise not be readily detectable. While an exhaustive list of the exact differences between phenotypes and aposematic signals would be rather hard to draw, we mention three of the most important differences here - the reader is referred to chapter 6 of Ruxton et al., (2019) for a more detailed discussion. First, aposematic signals are necessarily paired (at least in the mind of a predator) with some form of secondary defence (e.g. bright skin pigmentation is paired with unpalatable toxins). That is, the predator's response to the signal will be linked to a cognitive association the predator has formed between the signal and the defence or other aversive trait that makes prey unprofitable to predators. Second, they have evolved as signals through natural selection and are therefore effective in altering predator behaviour so that it favours prey survival (e.g. fewer recognition errors, enhanced wariness, accelerated learning and decelerated forgetting). Third, these alone can act as a primary defence mechanism in the sense that it is deterring for predators (through *learned aversion* - see Broom et al., 2006) and thus reduces the probability that an attack is mounted[2].

The observation of aposematism in the natural world would (upon first inspection) seem troubling from the evolutionary standpoint as it is sensible to surmise that conspicuous individuals run a clear disadvantage compared with their non-signalling relatives. Furthermore, there is considerable empirical evidence suggesting that chemical defences induce fitness costs (see the study by Dahl and Peckarsky, 2003 on chemically-defended fish showing *"a negative relationship between female allocation to eggs and to morphological defence character"* or those by Darst et al., 2006 or Zvereva and Kozlov, 2016 and those mentioned in the bibliography therein). Fitness costs are most likely associated with the costs of synthesising and/or acquiring (e.g. through dietary modifications) and/or storing toxins - see Daly (2003) or Darst et al. (2005). In several instances, these costs are directly manifested through reductions in growth and adult size, while in others, the cost can be calculated indirectly using energetic terms[3].

During the early stages of the evolution of aposematic defence prey may not have adapted completely to the defence on an organic/physiological level. For instance, in a study by Tarvin et al. (2017) poisonous *Dendrobatidae* frogs appeared to undergo a period of *"self-poisoning"*, which was notably less apparent among later-generation prey in which, perhaps due to observed ever-increasing amino acid replacements in toxin-binding sites. Aposematism is conceivably more interesting during these early stages because selective advantages associated with reduced predation are directly traded off against such "self-poisoning" (or equivalent) effects. In addition, chemical defences can be modelled as continuous trait characteristics and their frequent coupling with signalling cues that are external renders them especially interesting from the

[2]There are marginal cases (see Halpin et al., 2008) in which acoustic and/or visual signals are emitted by prey items exclusively during encounter. We do not consider these here.

[3]It should also be mentioned that there are cases in which there is no evidence of costs being incurred at all; such is the case with *Diprion pini* larvae discussed in Lindstedt et al. (2011) for example.

mathematical-modelling perspective. Indeed, much of the mathematical description provided here and in Broom et al. (2006) could best be thought to describe such systems (although it is by no means limited to this). Our description assumes that investment in toxicity is costly by involving a trade off in fecundity and is in this sense more early-stages focused.

It is argued that both predators and prey can benefit from honest signalling of chemical defences, particularly if there are costs to both parties associated with prey capture prior to detection of defences (such as time and energy invested in chasing and fleeing, and/or risk of injury). The review by Summers et al. (2015) reveals that there is a discrepancy both among empirical and among model-based studies relating to how conspicuousness and defence are related. One of the main objectives of Broom et al. (2006) had been to rephrase this question as *(a) how conspicuous and (b) how well-defended should a prey individual be so that a population made up entirely of one type can maintain their composition?* and to hence address this question within the context of an evolutionary model. We presently propose adaptations to the answers that had been provided therein by exploring a broader class of example functions (chapter 3). In the subsection that follows we review the first major contribution in this direction (Leimar et al., 1986), compare it with the model of Broom et al. (2006) and hence motivate the relevance of our own insight into this challenge.

The mathematical modelling of aposematism

To date, the co-evolution of defence and of the signalling of that defence (aposematism) has received limited theoretical attention and the systematic treatment of optimality in this context is also lacking compared with the study of defence in isolation. We begin our review of the mathematical treatments of aposematism with the celebrated work of Leimar et al. (1986), upon which the later work of Broom et al. (2005), Broom et al. (2006) and subsequently Broom et al. (2008) were developed. This work introduced complexity both in the behaviour of the predator (through varied rates of learning and attack probability) and also in the behaviour of prey, by allowing variations in the effectiveness of their strategies, which are measured in terms of their contribution to survival, their effects on predator learning rates and biological costs of defence and conspicuousness.

Leimar et al. (1986) considered structures in prey populations that emerge as a consequence of the assumptions described above and specify the patterns of ESS that are associated with these. The model of Leimar et al. (1986) predicts the possibility of a single cryptic ESS with non-trivial levels of defence and suggests that crypsis may be destabilised in favour of conspicuousness providing predation rates and kin grouping are sufficiently high. Moreover, it proposes that kin grouping is not necessary for the maintenance of aposematism but rather, a positive relationship between conspicuousness and learning is required together with increased avoidance of those more conspicuous than average phenotypes. In contrast to the model of Broom et al., 2006, the model of Leimar et al., 1986 considers a set of initially naive predators whose attack probabilities are highest before they have their first encounter ($G^{(0)}(x) = e(x)$, with x representing the level of prey unprofitability) and which continuously decrease over successive encounters according to an inhibitory gradient $h(x, x_1, y_1)$ (with x_1 being the unprofitability of encountered prey and y_1 its conspicuousness), so that

$$G^{(n)}(x, x_1, y_1) = e(x)[1 - h(x, x_1, y_1)]^n, \tag{2.1.1}$$

where n is the number of previous encounters[4]

[4]this is eq. (2) in the original publication; remark that we reserve the lowercase symbol g to describe a different quantity in the next section.

The model of Leimar et al. (1986) was the first to consider the joint action of aposematic traits and its accounting of the effects of predator behaviour on the evolution of aposematism is novel and self-contained. We should remark that equality (2.1.1) suggests that the attack probability is non-increasing with respect to n; in particular, it predicts that over successive encounters with prey (regardless of their conspicuousness, level of defence or the outcome of encounters) the predator's learned aversion is ever-increasing. This implies that the predator's consumption of prey is ever-decreasing and we contend that foraging strategies of this type would be unstable for predators in the long-run. Furthermore, the model was developed under the assumption of specific functional forms that make the results provided hard to generalise.

It should be noted at this stage that there are a number of studies in which it is incorrectly claimed that Leimar et al. (1986) suggest that optimal levels of conspicuous and defence are negatively correlated across populations. This is most recently done in Summers et al. (2015), which constitutes an otherwise indispensable review of empirical and model-based approaches on honest signalling and indeed one that we consult in our discussions in later chapters. As it happens, the model of Leimar et al. (1986) does not consider signal strength at all, it only compares a non-signalling (camouflaged) phenotype to a signalling phenotype. In particular, it is argued that if the signalling phenotype is associated with a reduced rate of attack by predators (perhaps through learned aversion) such that the non-predatory cost of producing toxins increases with its effectiveness in reducing predator attack rates then the optimal strategy may be for the signalling phenotype to invest less in costly toxins than the non-signalling phenotype. The possibility of a true, negative correlation between signal strength and level of defence within a causal game-theoretical framework is conceived in Scaramangas and Broom (2022) and forms a novel extension to the predictions of Broom et al. (2006). We return to this point in chapter 3.

There are a number of key considerations that are different between this presentation and that of Leimar et al. (1986), which make the two approaches complementary to each other. For instance in their model, predators are initially naive and through repeated experiences engage in aversive learning, which changes the level of predation pressure over time. We contend that such a description would be especially purposeful for studying the initial evolution of aposematism in a set of seasonal predators who arise inexperienced at the start of the season (see Kauppinen and Mappes, 2003 for an example of a dragonfly species *Aeshna grandis* or *Brown hawker* who typically fly between the months of July and September and in this period learn to avoid aposematic wasps *Vespula norwegica*). In this instance, predators are short lived compared to the time taken to learn about the prey population, such that the process of learning is important to overall prey mortality.

By contrast, the model of Broom et al. (2006) describes an equilibrium situation among prey-predator populations in which aposematism has already become established and such that there is no change in predation pressure over time. That is, the equilibrium in Broom et al. (2006) refers to average rates of learning, hunger, experience, age, etc. among predators and is perhaps maintained through a balanced mix of young and old individuals in overlapping generations. We should remark that while equilibrium is assumed, how it is reached (presumably through aversive learning and/or genetic inheritance and/or strong transmission of foraging behaviours among kin groups) is not explicitly a feature of the model. Implicit is the assumption that the process of learning involves only a short fraction of the predator's lifetime and the overwhelming majority of the mortality that the prey population experiences is caused by experienced predators that have completed their learning and the process of learning itself can be ignored from the prey perspective.

In Broom et al. (2006) we establish that while learning is important it is assumed that predators learn

quickly, so that most of their life they impose morality based on their understanding of prey traits gathered during a short investigative learning phase early in life. The type of system that this model might best describe are insectivorous birds – these birds might eat hundreds of insects a day and live for several years. If the prey population is consistent under that timescale, which might suggest a tropical rather than temperate region (putative predators could include the *Greater rhea* or species of heron, although this geographical restriction is not a hard one) then we might expect that birds can learn about prey that they readily encounter on a timescale of days which is much shorter than their lifetime. For instance, there is ample evidence to suggest that the diet of the American robin (*Turdus migratorius*) includes caterpillars - see capter 2 of Martin et al. (1961) and cuckoos in particular are known to attack even the most unpalatable ones - see Barbaro and Battisti (2011).

The theoretical implications of Broom et al. (2006) are developed further in Broom et al. (2008). Both papers consider non-point solutions and the reader is encouraged to consult those which are presently mostly omitted. It is worth noting that many of the more recent mathematical models focus on the signalling component of aposematism, its effectiveness in promoting aversive learning (see for example Summers et al., 2015, Merilaita and Ruxton, 2007, Broom et al., 2013). Other works focus more heavily on the component of defence and on the costs that these incur on the prey that acquire them (see for example L. Wang and Broom, 2019, or Broom et al., 2010). Other important work aims at underpinning the connections between static and the dynamical approaches in evolutionary games (see Argasinski and Broom, 2018 or Argasinski, 2006). Notable is also the vast array of numerical and simulation techniques that have been implemented and which have provided unique insight into the evolutionary process (see Teichmann et al., 2015,Teichmann et al., 2014a,M. P. Speed et al., 2010, G. D. Ruxton et al., 2009, or M. P. Speed and Ruxton, 2007).

2.2 Model description

In this section we describe a more generalised version of the model by Broom et al. (2006) by accounting for a non-negligible proportion of mutant prey and arriving at the description presented in Broom et al. (2006) by taking the limit as this tends to zero (see also Scaramangas and Broom, 2022). We begin with a generalised description (which we use in chapter 4) before proceeding the resident-mutant description, which is used in the following section to derive the conditions for evolutionary stability. We should remark that a number of novel clarifications are made in this section that (hopefully) admit an improvement to the original presentation in Broom et al. (2006). The majority of these improvements are also seen in Scaramangas and Broom (2022) and in Scaramangas et al. (2022).

Unrestricted prey strategies

We begin by considering a population of prey of a certain species who occupy some habitat. Assume that the habitat can be partitioned into effectively infinite, non-overlapping localities consisting of approximately N prey, where N is taken to be large. In addition, assume that each locality is visited by n predators, who visit a single locality only. Implicit in this layout is that there is a uniform territorial division of the habitat among the predators so that each locality can be perceived as the territory of a specific number of n predators (n is too taken to be large). It is also generally assumed that the overall population of prey and predators are in dynamic equilibrium (predators are on average characterised by fixed states of hunger level, age, experience etc. - see earlier discussion) so that in any one locality predation has no effect on the relative sizes N and n. Within each locality prey are labelled by index i such that $\{i = 1, 2, ..., N\}$.

Prey defend themselves against predators by investing in secondary defences of strength $t_i \geq 0$, which they advertise through recognisable signals of *conspicuousness* $r_i \geq 0$; naturally completely inconspicuous (or *cryptic*) prey have $r_i = 0$. In addition to the possession of a conspicuous signalling trait prey assume a *colouration* trait $\theta_i \in [0, 2\pi)$, which is distinct to the conspicuousness. The assumption that these traits are distinct (most plausible for chemically-defended prey[5] whose levels of unpalatability, pigmentation intensity and colouration are separate - see earlier discussion on signalling cues that are *external*) is for the purposes of mathematical modelling synonymous with the assumption that these can be varied individually and independently from one another and can therefore represented by the 3-vector (r_i, t_i, θ_i), which denotes the *strategy* of prey individual i in a cylindrical coordinate system. It is worth mentioning that while this 3-vector description is used in Broom et al. (2006) we suppress this third variable in the chapters that follow and continue the remainder of the discussion by representing prey strategies as vectors $(r_i, t_i) \in \mathbb{R}^{\geq 0} \times \mathbb{R}^{\geq 0}$. In addition, it is assumed that individual i reproduces with *fecundity* rate $F = F(t_i)$, where F is a declining function of t_i (indicating that investment in toxins is costly), dies of causes other than predation at some fixed *background mortality rate* λ and defend themselves against predators by acquiring aposematic traits (r_i, t_i).

It has been mentioned that the model of Broom et al. (2006) is static and does not impose specific dynamics to describe prey-predator interactions. We surmise that on average predators encounter prey at some fixed rate σ. Detection of individual i is an event that is conditional on the predator encountering that prey such that the rate of detection $D(r_i)$ can be defined as the product of this fixed rate of encounter and the probability that i is detected given encounter has occurred. We write

$$\text{Rate of detection of } i = \sigma \times \mathbb{P}(i \text{ is detected} \,|\, i \text{ is encountered}) \tag{2.2.1}$$

Upon detecting prey, predators can decide to mount an attack, which may or may not lead to capture. We express the *detection rate* as $D(r_i)$, which we assume to be an increasing function of r_i, which tends to unity as prey conspicuousness assumes arbitrarily large values - indeed an interval of unit time can be defined as the time taken for a very bright prey item to be detected by a predator - while $D(0) = d_0 > 0$, since even fully-cryptic prey can be detected. The probability that a mounted attack results in capture is given by $K = K(t_i)$, where K is declining with t_i, indicating that more defended individuals are harder to capture. A detected individual will be attacked with probability $Q = Q(I_i)$, where I_i represents the *average aversive information* that the average predator has on item i. It is assumed that Q is a declining function of I_i and such that $Q(I_i) = 1$ for $I_i \ll -1$ and $Q(I_i) \approx 0$ for $I_i \gg 1$; these assumptions suggest that a prey individual perceived by the predator as very attractive ($I_i \approx 0$) are most-likely to be attacked, while those perceived as very aversive ($I_i \gg 1$) are unlikely to be attacked. We return to this point once we have discussed the significance of the perceived aversiveness.

Predators assign I_i to individual i by comparing it to a certain (weighted) base-line level of aversive information, which is generated through encounters with the prey population, such that

$$I_i = \frac{1}{n} \sum_{j=1, j\neq i}^{N} L(r_j) H(t_j) \mathcal{S}(r_i, r_j). \tag{2.2.2}$$

It is understood that each locality is sizeable (or rather, it can occupy a large number of prey individuals such

[5] A robust example can be drawn from species of the Poison dart frog (*Dendrobatidae*) family who signal chemical defences through bright skin pigmentation with an array of colours. In our description toxicity, brightness and colouration are treated as traits that are independent and fully identify them as aposematic.

that N can be taken to be large) and is visited by a group of predators n, who visit this locality only. Tacit in this discussion is that it is (usually) not possible for any one predator to experience every single prey item within its life-cycle; rather, we assume that predators experience the locality collectively and the aversiveness of their experiences is shared equally among them (see factor of $1/n$ in (2.2.2)). Even though as a collective, predators have complete experience of the locality, their perceived aversiveness of a particular individual i is not drawn directly by their experience of it, but through successive experiences with its neighbours (notice that the sum in (2.2.2) excludes i, but includes the remaining locality).

An important assumption of our model is that predators learn quickly, so that most of their life they impose mortality based on their understanding of prey traits gathered during a short investigative learning phase early in life. Since learning involves only a short fraction of the predator's lifetime it can be mostly ignored from the prey perspective. Indeed, it is assumed that the overwhelming majority of prey mortality is caused by experienced predators, who have completed the learning process described in (2.2.2) - see discussion in the latter part of section 2.

Predators find chemically-defended prey aversive and the experience of consuming them is measured by $H = H(t_i)$, which is an increasing function of t_i and is zeroed at a critical value of the toxicity $t_i = t_c$. We write

$$H(t_i) \begin{cases} < 0, & t_i < t_c \\ = 0, & t_i = t_c \\ > 0, & t_i > t_c. \end{cases} \tag{2.2.3}$$

The level of defence of prey with $t_i < t_c$ is not sufficient to outweigh the nutritional benefit received from predators by consuming them and are perceived as attractive or *negatively aversive*. By construction, the defence of prey with $t_i = t_c$ describes the limiting value at which the nutritional benefit is exactly outweighed by their distastefulness and such prey are perceived as *neutrally aversive*. Finally, prey with $t_i > t_c$ are perceived as *(positively) aversive* by the potential predator. The second term on the RHS of (2.2.2) requiring explanation is $L = L(r_i)$, which represents the rate at which encounters that have occurred can be recalled by predators. This is assumed to be a growing function of r_i indicating that encounters with more conspicuous prey can be better recalled. In much of the later work we will assume *perfect predator recollection*, which involves taking taking $L = D$.

The third term that warrants explanation is that which appears on the RHS of (2.2.2) and is the *similarity function*

$$\mathcal{S} : \mathbb{R}^{\geq 0} \times \mathbb{R}^{\geq 0} \to [0, 1] \text{ such that } (r_i, r_j) \mapsto \mathcal{S}(r_i, r_j), \tag{2.2.4}$$

which describes how predators perceive the *visual similarity* of different-looking prey. In contrast, the (generalised) similarity function $\mathcal{S}$ of (2.2.4) - denoted with calligraphy is naturally bi-variate. In Broom et al. (2006) and indeed for the remainder of this report we treat this as a a uni-variate function S of the Euclidean distance separating their visual appearance. In particular we impose that

$$\mathcal{S}(r_i, r_j) = S(|r_i - r_j|), \quad \text{where } S : \mathbb{R} \to [0, 1] \text{ is s.t. } x \mapsto S(x). \tag{2.2.5}$$

We should remark that it is only in Scaramangas and Broom (2022) that the distinction between the uni-variate and the bi-variate functions associated with perceived similarity as in (2.2.5) is drawn. The uni-variate function S on the RHS of (2.2.5) is assumed to be C^l with $l \geq 2$ (at least sufficiently near the origin) and

has the following properties

$$[i]\ S(x) \in [0,1] \text{ for all } x \geq 0 \tag{2.2.6}$$

$$[ii]\ S(0) = 1 \ - \ \text{i.e. } S(|r_i - r_j|) = 1 \text{ if and only if } i \text{ and } j \text{ are equally conspicuous with } r_i = r_j \tag{2.2.7}$$

$$[iii]\ \lim_{x\to\infty} S(x) = 0 \ - \ \text{i.e. } S(|r_i - r_j|) = 0 \text{ iff } i \text{ and } j \text{ are very dissimilar, s.t. } r_i \gg r_j \text{ or } r_j \gg r_i \tag{2.2.8}$$

$$[iv]\ S'(x) \leq 0 \text{ for all } x > 0 \ - \ \text{i.e. } S \text{ does not increases as } |r_i - r_j| \text{ increases} \tag{2.2.9}$$

$$[v]\ S'(0) < 0 \ - \ \text{i.e. variations in conspicuousness (incremental) are perceptible at the baseline} \tag{2.2.10}$$

$$[vi]\ S''(x) \geq 0 \ - \ \text{i.e. predator sensitivity to incremental variations is non-increasing} \tag{2.2.11}$$

While properties $[i], [ii]$ and $[iii]$ are a matter of definition, properties $[iv]$, $[v]$ and $[vi]$ are dependent on our underlying assumptions about predator psychology. We deliberate on this last point presently. In order to do so, we may, for the time being (and without loss of generality in any of the claims that follow) imagine that predators are (on average) used to encountering individuals resembling i, so that r_i represents some baseline level of appearances. In addition, we consider individual j that is more conspicuous than i (i.e. $0 < r_i < r_j$) such that $|r_i - r_j| =: x_*$, where x_* is sufficiently near the origin (so that condition S being $\mathcal{C}^l$ with $l \geq 2$ holds - see later discussion). Incremental variations in r_i and r_j are described by quantities $r \in [r_i - \delta r, r_i + \delta r]$ and $\hat{r} \in [r_j - \delta r, r_j + \delta r]$, where $0 < \delta r \ll 1$ and can be used to express the derivative with respect to x at x^* as

$$S'(x_*) = \lim_{x\to x_*} \frac{S(x) - S(x_*)}{x - x_*} = \lim_{\hat{r}\to r_j} \frac{S(|r_i - r_j|) - S(|r_i - \hat{r}|)}{|\hat{r} - r_j|}. \tag{2.2.12}$$

This quantity captures the rate at which the average predator perceives incremental variations in the visual appearance of prey at some "distance" x_* away from the baseline. A basic reading of condition $[iv]$ is that as this distance x away from the baseline increases the elevation of S does not increase. Namely that if i and j are assigned some level of similarity $S(x_*)$ then the associated level of S corresponding to a more conspicuous individual playing $\hat{r} \in (r_j, r_j + \delta r]$ is not larger. Indeed, condition $[iv]$ guarantees that a first-order expansion about x_* in this direction[6] $S(|r_i - r_j|) + |\hat{r} - r_j| \times S'(x_*) \leq S(x_*)$, as required. It also follows that small differences from the baseline itself can be determined by evaluating the derivative $S'(x)$ at $x = 0$. That is

$$S'(0) = \lim_{r\to r_i} \frac{S(|r - r_i|) - S(|r_i - r_i|)}{|r - r_i|}. \tag{2.2.13}$$

Possible violations of condition $[v]$ would include cases with $S'(0) > 0$ or $S'(0) = 0$. The former is immediately rejected as it violates requirement $[i]$ - indeed $S(|r - r_i|) \approx S(|r_i - r_i|) + |r - r_i| \times S'(0) > 1$. As for the possibility $S'(0) = 0$ - examples could include Gaussian forms on $x \geq 0$ (discussed below) - we remark the following. Since S is bounded from above and from below (condition $[i]$), it is non-increasing within these bounds (condition $[iv]$) and approaches the lower bound for large enough x (condition $[iii]$) it follows that j could be chosen so that $S'(x_*) < 0$, once more maintaining the requirement that S is $\mathcal{C}^l$ with $l \geq 2$ at this value. Without loss of generality, we assume that $r \in (r_i, r_i + \delta r]$ and $\hat{r} \in (r_j, r_j + \delta r]$ so that

[6] Expansion in the opposite direction can be achieved by setting $\hat{r} \in [r_j - \delta r, r_j)$ and yields the reverse inequality, namely $S(|r_i - r_j|) - |\hat{r} - r_j| \times S'(x_*) \geq S(x_*)$

from (2.2.12) it now follows that the change in elevation of S at such an $x = x_*$ can be approximated by $S(|\hat{r} - r_i|) - S(|r_i - r_j|) \approx |\hat{r} - r_j| \times S'(0) < 0$. From (2.2.13) and the assumption that $S'(0) = 0$ it also follows that the elevation of S does not change at $x = 0$ since $S(|r - r_i|) - S(|r_i - r_i|) \approx |r - r_i| \times S'(0) = 0$.

We have demonstrated that $S'(0) = 0$ is the only potential alternative to $[v]$ and that if this were to hold it would imply that predators are (on average) more sensitive to variations in appearance when these occur far from the baseline but not at the baseline itself. Such a conclusion seems to suggest that predators can distinguish small changes in the appearance of prey types that they are not used to encountering but not in the types that they are used to encountering. We might expect that such a result is less relevant for keen-sighted avian predators feeding on Poison dart frogs, to which the model of Broom et al. (2006) is adept (but not limited) to describing. Throughout this manuscript we insist on condition $[v]$ and exclude similarity functions that are flat-peaked at the origin from our discussions. The reader is encouraged to consult Balogh and Leimar (2005) for an illustration of the use of flat-peaked generalisation curves - this is done in the context of the evolution of mimicry - and a discussion of restrictions on the shapes that these can assume.

We have discussed how the derivative in (2.2.12) describes the predator's sensitivity to small changes in the visual appearance of prey at some arbitrary distance away from the baseline. It follows that if the reverse of condition $[vi]$ were to hold (i.e. $S''(x) < 0$) that predators would become increasingly sensitive to small variations as this distance increased. This is not a consequence we wish to entertain and deems condition $[vi]$ necessary. Requiring $[vi]$ is to say that predators gradually become less sensitive (or remain equally sensitive) to these variations as the distance away from the baseline increases. In all worked examples we consider in this manuscript, we make use of the function $S(x) = \max(1 - vx, x)$ for which $S''(x) = 0$ for all $x \geq 0$[7]. With such a generalisation the predator is equally sensitive to small variations at all distances away from the baseline (this also applies to individuals that are perceived as completely dissimilar with $x > 1/v$). We argue that such a function is favourable for modelling purposes as it imposes no specific condition on how predator sensitivity varies away from the baseline.

If we did not place this level of importance on condition $[v]$ perhaps the natural choice for us to consider would be a Gaussian similarity function of the form $S(x) = \exp(-x^2/\varepsilon)$. Despite being flat-peaked at the origin, allowing parameter $\varepsilon > 0$ to assume arbitrarily small values enables us to think of the predator as being sensitive to small variations in appearance near (though not exactly at) the baseline. While for the reasons mentioned this is an example worth considering (and we would if time resources were ample), it is also worth noting that it violates condition $[vi]$. Indeed, $S''(x) < 0$ on $[0, \sqrt{\varepsilon/2})$ so that on this interval predator sensitivity drops rapidly ($S'(x) \sim 1/\varepsilon$ with $\varepsilon \ll 1$) away from the baseline. This is is not an effect we consider plausible and in addition, the function imposes specific requirements on the behaviour of predator sensitivity away from the baseline. Indeed, $S'(x)$ admits a local minimum at $x = \sqrt{\varepsilon/2}$ so that the predator generalises at an (absolute) equal rate pairs of points that are at vastly different distances away from the origin. For instance for $\varepsilon = 0.01$ points $x \approx 5.01 \times 10^{-4}$ and $x \approx 2.492 \times 10^{-1}$, differing by three order of magnitude admit the same value of predator sensitivity).

The rate at which predators encounter prey occurs on average with some fixed rate σ. An encountered prey may or may not be detected (see (2.2.1)) and it is assumed that the more conspicuous the encountered prey is, the likelier it is for this to be detected (although most plausible functional forms for $D(r_i)$ should exhibit some plateau indicating diminishing impact of ever-increasing conspicuousness on the rate of detection).

[7] As we establish the exclusion of $x = 1/v$ from this claim (the derivative of $S(x) = \max(1 - vx, 0)$ is not defined at $x = 1/v$) is of limited importance for exploring evolutionary stability

Symbol	Meaning
r	the conspicuousness of a prey individual
t	the toxicity of a prey individual
N	the size of the prey population
n	the size of the predator population
$D(r)$	the rate at which r-individuals are detected
$L(r)$	the rate at which r-individuals are detected and recalled
$\mathcal{S}(r_i, r_j)$	the generalised (bi-variate) similarity between individuals with conspicuousness r_i and r_j
$S(x)$	the uni-variate similarity function of individuals differing in appearance by x
$H(t)$	the aversiveness of prey individuals with toxicity t
t_c	the critical level of toxicity such that $H(t_c) = 0$
$F(t)$	the fecundity of a prey individual with toxicity t
$K(t)$	the probability that an attacked t-individual is captured
$Q(I)$	the probability that a detected I-individual is attacked
I	the level of aversive information of an individual
λ	the prey background mortality rate (not due to predation)
a	the average relatedness of prey individuals in the population

Table 2.1: The parameters and functions of the model.

Detected prey may or may not be attacked and for prey who are known to be aversive this is less likely - see $Q(I_i)$ and explanations of (2.2.2). Finally, attacked prey may or may not be captured and for better defended prey this is less likely. We can determine the rate at which prey are captured by predators by observing that for any individual i capture is conditional on attack, which is conditional on detection, which is in turn conditional on encounter. We write

$$\begin{aligned}\text{Rate that } i \text{ is captured} =& \sigma \times \mathbb{P}(i \text{ is detected} \,|\, i \text{ is encountered}) \\ & \times \mathbb{P}(i \text{ is attacked} \,|\, i \text{ is detected}) \times \mathbb{P}(i \text{ is captured} \,|\, i \text{ is attacked}) \\ =& D(r_i)K(t_i)Q(I_i)\end{aligned}$$

to describe the *predator-induced mortality rate of* i in terms of the functional forms of Table 2.1. For the *total mortality rate of* i we write

$$\text{Rate that } i \text{ is killed} = \lambda + D(r_i)K(t_i)Q(I_i). \tag{2.2.14}$$

We should remark that since the mortality rate is naturally in units of inverse time, the reciprocal of the total mortality rate (in units of time) describes the average life-cycle of i. Since the fecundity rate $F(t_i)$ describes the number of offspring produced by i in units of inverse time it follows that

$$P(r_i, t_i) = \frac{F(t_i)}{\lambda + D(r_i)K(t_i)Q(I_i)} \tag{2.2.15}$$

is a unitless quantity describing the number of offspring produced per life-cycle by individual i and is therefore a measure of its *fitness*. We have reserved the letter P to identify the quantity on the RHS of (2.2.15) with the *payoff to i*. The payoff (or fitness) is perhaps the most central of this manuscript and we will continue to refer back to this throughout. The presentation of the model of Broom et al. (2006) has been kept general; in the part that follows we will assume a specific structure for the prey population and focus on combinations of the defensive and signalling trait that are evolutionarily stable.

Residents vs. mutants

In keeping with the focus of Broom et al. (2006) we consider a resident-mutant prey setup in order to study the evolutionary stability of aposematic traits. That is, we imagine that the majority of prey play some *resident strategy* (r_1, t_1), while a much smaller fraction $\varepsilon \ll 1$ play some *mutant strategy* (r, t). Mutations in the genomes associated with aposematic phenotypes are rare and are typically manifest as small errors and from the mathematical modelling perspective we assume that the mutant strategy is *local* to the resident strategy. It is immediately obvious that the local vicinity of a resident strategy depends on where in the strategy space the resident strategy is drawn from. We identify four sets ($\mathcal{D}_0$, $\mathcal{D}_1$, $\mathcal{D}_2$ and $\mathcal{D}_3$) as making up the strategy space $\mathcal{D}$ such that

$$\underbrace{\{\rho \geq 0, \tau \geq 0\}}_{\mathcal{D}} = \underbrace{\{(0,0)\}}_{\mathcal{D}_0} \sqcup \underbrace{\{\rho > 0, \tau = 0\}}_{\mathcal{D}_1} \sqcup \underbrace{(\rho = 0, \tau > 0)}_{\mathcal{D}_2} \sqcup \underbrace{\{\rho > 0, \tau > 0\}}_{\mathcal{D}_3}. \tag{2.2.16}$$

Assuming that quantities δr and δt are positive and arbitrarily small (i.e. $0 < \delta r \ll 1$ and $0 < \delta t \ll 1$) we distinguish the following cases of local neighbourhoods for the mutant strategy. We have that

$$\text{if } (r_1, t_1) \in \mathcal{D}_0 \text{ then } (r, t) \in [0, \delta r] \times [0, \delta t], \tag{2.2.17}$$

$$\text{if } (r_1, t_1) \in \mathcal{D}_1 \text{ then } (r, t) \in [r_1 - \delta r, r_1 + \delta r] \times [0, \delta t], \tag{2.2.18}$$

$$\text{if } (r_1, t_1) \in \mathcal{D}_2 \text{ then } (r, t) \in [0, \delta r] \times [t_1 - \delta t, t_1 + \delta t] \text{ and} \tag{2.2.19}$$

$$\text{if } (r_1, t_1) \in \mathcal{D}_3 \text{ then } (r, t) \in [r_1 - \delta r, r_1 + \delta r] \times [t_1 - \delta t, t_1 + \delta t]. \tag{2.2.20}$$

As we discover in the section that follows, distinguishing between the types of neighbourhood as in (2.2.17) through to (2.2.20) is key to determining the conditions for evolutionary stability that prevail in these (as it happens, in the $\varepsilon \to 0$ limit, these are distinct).

Before doing so, we establish a number of clarifications that describe the model of Broom et al. (2006) in a more general setting. We now consider a certain *focal individual*, whose strategy (ρ, τ) we leave unspecified by setting $(\rho, \tau) \in \{(r, t), (r_1, t_1)\}$. To that, we introduce imperfect mixing to the prey population by suggesting that in a small, finite number of localities there are *relatives* of the focal individual. We assume that these relatives are perfect copies of the focal individual and form *mutant colonies* that are assumed to be rare. Examples of colonies are seen in a number of different populations. In Cole (1946) the phenomenon of *"clumping of individuals into groups"* is described such that *each group may be relatively or entirely independent of all similar groups and, therefore, that these distributional units may be randomly distributed"*. A more elaborate discussion of the spatial distribution of insect populations can be found in Taylor (1984). Interestingly, examples of amphibian populations, whose chemical defences are in keeping with the description

also form colonies. For instance, this is the case for *Polypedates leucomystax* frogs that are examined in Roy, 1997).

Imperfect mixing is introduced by imposing local *clustering* on an otherwise well-mixed population, which simply means that the prey population is mostly well-mixed except for certain regions of the habitat where this is not the case. In practice, we imagine that in almost all sites of the habitat there is a fixed proportion of εN mutants and $(1-\varepsilon)N$ residents except for a small number of these in which this proportion is perturbed (either there are slightly more residents than usual or slightly more mutants than usual). The extent to which the proportion is different in these localities is controlled by the *local relatedness parameter* $a \in [0,1]$, so that when $a = 0$ there is no impact on the background mutant/resident proportion and we recover the well-mixed regime (see (2.2.25) and (2.2.26)). If $a = 1$ this corresponds to a scenario in which the localities are made up entirely of one or the other type. More generally, if a is non-zero we expect that there are $aN - 1$ relatives making up the colony (excluding the focal individual) and $(1-a)N$ non-relatives of which $(1-a)\varepsilon N$ play the mutant strategy and $(1-a)(1-\varepsilon)N$ play the resident strategy. Parameter a is interpreted as a quantity that measures the concentration of relatives in the local area and implicit in this is that the focal individual breeds true.

By introducing a suitable partitioning $\{1,...,N-1\} = T_0 \sqcup T_1 \sqcup T_2$, with $T_0 = \{1,...,aN-1\}$, $T_1 = \{aN,...,(1-a)\varepsilon N + aN - 1\}$ and $T_2 = \{(1-a)\varepsilon N + aN,...,N-1\}$ and labelling with $\mathcal{I}^\varepsilon(\rho,\tau;r,t;r_1,t_1)$ the perceived aversiveness of the focal individual it is clear that through (2.2.2) this amounts to

$$\begin{aligned}\mathcal{I}^\varepsilon(\rho,\tau\,;\,r,t\,;\,r_1,t_1) &= \frac{1}{n}\sum_{j\in T_0} L(r_j)H(t_j)S(|\rho - r_j|) + \frac{1}{n}\sum_{j\in T_1} L(r_j)H(t_j)S(|\rho - r_j|)\\ &\quad + \frac{1}{n}\sum_{j\in T_2} L(r_j)H(t_j)S(|\rho - r_j|)\\ &= \frac{1}{n}(aN-1)L(\rho)H(\tau)S(|\rho-\rho|) + (1-a)\varepsilon\frac{N}{n}L(r)H(t)S(|\rho - r|)\\ &\quad + (1-a)(1-\varepsilon)\frac{N}{n}L(r_1)H(t_1)S(|\rho - r_1|). \end{aligned} \tag{2.2.21}$$

We emphasize that parameters ε and a are distinct and describe fundamentally different quantities. While the former describes the background proportion of mutants over the entire prey habitat the latter describes the proportion of focal relatives in a (rare) site in which clustering is present. Implicit in (2.2.21) is that we are working is such a (rare) site. In addition we remark that if most localities can hold a large enough number of prey (i.e. $N \gg 1$) it is safe to assume for most reasonable values of the local relatedness a that the inclusion or removal of the focal individual from within its group of relatives will not have a notable impact on its perceived aversiveness so that we can take $(aN-1)/n \approx aN/n$ on the RHS of (2.2.21). So now the aversiveness of the focal individual is provided by the slightly simpler expression

$$\mathcal{I}^\varepsilon(\rho,\tau\,;\,r,t;\,r_1,t_1\,) \approx a\frac{N}{n}L(\rho)H(\tau) + (1-a)\varepsilon\frac{N}{n}L(r)H(t)S(|\rho-r|) + (1-a)(1-\varepsilon)\frac{N}{n}L(r_1)H(t_1)S(|\rho-r_1|), \tag{2.2.22}$$

which we use to generate the aversiveness of the mutant and the resident as follows. If the focal individual is a mutant this means that there are $aN + (1-a)\varepsilon N$ mutants and $(1-a)(1-\varepsilon)N$ residents present in the mentioned locality. Setting $(\rho,\tau) = (r,t)$ into (2.2.22) gives us the mutant aversiveness (denoted in non-calligraphy), through the focal expression as $I^\varepsilon(r,t;r_1,t_1) := \mathcal{I}^\varepsilon(\rho = r, \tau = t; r,t;r_1,t_1)$. Because this

quantity is determining for the probability of attack of the mutant type and features as an argument in Q - see (2.2.15) - we tend to suppress its argument and use the shorthand $I^\varepsilon \leftrightarrow I^\varepsilon(r,t;r_1,t_1)$. This quantity is given by

$$I^\varepsilon = a\frac{N}{n}L(r)H(t) + (1-a)\varepsilon\frac{N}{n}L(r)H(t) + (1-a)(1-\varepsilon)\frac{N}{n}L(r_1)H(t_1)S(|r-r_1|). \tag{2.2.23}$$

Similarly, if the focal individual is a resident this means that there are $aN + (1-a)(1-\varepsilon)N$ residents and $(1-a)\varepsilon N$ mutant prey in the locality. The resident aversiveness (denoted in non-calligraphy) is given through the focal aversiveness in (2.2.22) as $I_1^\varepsilon(r_1,t_1;r,t) := \mathcal{I}^\varepsilon(\rho = r_1, \tau = t_1; r_1, t_1; r, t)$. As with the case of the mutant, we tend to suppress the arguments of the resident aversiveness to write $I_1^\varepsilon(r_1,t_1;r,t) \leftrightarrow I_1^\varepsilon$. This quantity reads

$$I_1^\varepsilon = a\frac{N}{n}L(r_1)H(t_1) + (1-a)\varepsilon\frac{N}{n}L(r)H(t)S(|r-r_1|) + (1-a)(1-\varepsilon)\frac{N}{n}L(r_1)H(t_1). \tag{2.2.24}$$

It was mentioned previously that the measures of mutant and resident aversiveness (as given through (2.2.23) and (2.2.24)) are evaluated in sites on which clustering is present. Imperfect mixing is realised by imposing rare clustering in an otherwise well-mixed population of prey. The associated levels of mutant and resident aversivess evaluated as averages would correspond to an effectively well-mixed population. In practice, this is recovered by considering the limit in which the local relatedness parameter tends to zero ($a \to 0$). Taking $a \to 0$ by definition amounts to considering a rare site in which the background proportion of mutants and residents is perturbed, but the extent to which it is is negligible (the focal individual is in among too few relatives). This picture is equivalent to perfect mixing, which is turn representative of the average predator's experience. Their perceived aversiveness is evaluated by taking the $a \to 0$ limit of expressions (2.2.23) and (2.2.24), which amounts to

$$\lim_{a\to 0} I^\varepsilon = \varepsilon\frac{N}{n}L(r)H(t) + (1-\varepsilon)\frac{N}{n}L(r_1)H(t_1)S(|r-r_1|) \tag{2.2.25}$$

and

$$\lim_{a\to 0} I_1^\varepsilon = (1-\varepsilon)\frac{N}{n}L(r_1)H(t_1) + \varepsilon\frac{N}{n}L(r)H(t)S(|r-r_1|). \tag{2.2.26}$$

So far we have discussed the significance of parameters a and ε in the context of the habitat structure and assumed that the mutant traits are fixed and given by (r,t). For consistency, we mention that if there is no mutation present, i.e. if $(r,t) = (r_1,t_1)$ that we recover the scenario we would expect. Indeed, setting $(\rho,\tau) = (r,t) = (r_1,t_1)$ into (2.2.22) gives

$$\mathcal{I}^\varepsilon(\rho = r_1, \tau = t_1; r = r_1, t = t_1; r_1, t_1) = \frac{N-1}{n}L(r_1)H(t_1) \approx \frac{N}{n}L(r_1)H(t_1), \tag{2.2.27}$$

which corresponds to the aversiveness of prey as perceived by predators visiting a habitat made up entirely of residents.

In closing, we introduce the notion of payoff. From expressions (2.2.23) and (2.2.24) it is clear that specifying the mutant and resident strategies together with their relative abundances specifies their perceived aversiveness; in light of (2.2.15) this also specifies their fitness. The fitness of the mutant reads

$$P^\varepsilon(r,t;r_1,t_1) = \frac{F(t)}{\lambda + D(r)K(t)Q(I^\varepsilon)}, \tag{2.2.28}$$

while the fitness of the resident reads

$$P_1^{\varepsilon}(r_1, t_1; r_1, t_1) = \frac{F(t_1)}{\lambda + K(t_1)D(r_1)Q(I_1^{\varepsilon})}. \tag{2.2.29}$$

Since both I^{ε} and I_1^{ε} as given in (2.2.23) and (2.2.24) evaluated with $(r, t) = (r_1, t_1)$ correspond to the common expression (2.2.27) it follows that $P(r = r_1, t = t_1; r_1, t_1) = P_1(r_1, t_1; r_1, t_1)$, which is consistent with the intuition that the mutant fitness matches the resident fitness precisely when the mutant traits match the resident value.

So far in this section we have re-introduced the model of Broom et al. (2006) by incorporating into this the amendments that are present in Scaramangas and Broom (2022). Indeed, in this section can be found the more up-to-date (and accurate) discussions of Scaramangas and Broom (2022) relating to the territorial-division of the habitat, the concept of local relatedness in the more generalised setting of mutation that includes a finite proportion of mutants ε. In this section we have also distinguished between the bi-variate similarity function and the uni-variate function through (2.2.5), which are found under one common (and slightly misleading) symbol in Broom et al. (2006). In the section that follows we demonstrate how the conditions for local evolutionary stability provided in Broom et al. (2006) correspond to the $\varepsilon \to 0$ regime introduced in this section. The joint consideration of residents with mutants in negligible proportions best describes the initial evolution of mutation and is the correct setting in which to study the evolutionary stability of aposematic traits. Indeed, if the residents can withstand invasion against mutants that have strong presence on the local level (controlled by a) but still an insignificant presence overall ($\varepsilon \to 0$) then this is sufficient to guarantee the evolutionary stability of the residents.

2.3 The evolutionary stability of aposematism

It is of particular interest to consider the resident-mutant representation (introduced in the latter part of the previous section) in the limit as $\varepsilon \to 0$, in which the overall proportion of mutants in the prey population is effectively zero. At first sight it may seem like a static model that considers a prey population in which everyone plays the same strategy is of no interest. Indeed, the contrary is true. It is possible to consider the limit in which $\varepsilon \to 0$ and still maintain that mutation is present. From the mathematical modelling perspective, this is achieved in two ways. First, by allowing potentially large clusters of focal relatives (the focal individual may be a mutant - see earlier discussion) to be present scarcely in among the habitat patches in a way that their contribution to the population inhabiting the entire habitat structure is negligible. Second, by maintaining that the habitat structure is territorially-divided among predators so that each site/patch is inhabited by N prey and visited by n predators who visit that patch only.

In the $\varepsilon \to 0$ limit predators visiting a site containing a cluster of mutants will have a different level of perceived aversiveness compared with the average predator who visits sites consisting only of residents. To be specific, the sign and magnitude of this difference will depend on the resident strategy being played, the mutant strategy being played and on the abundance of mutants in the given site. It is this territorial division of the habitat among predators that allows us to account for differences in the perceived aversiveness of mutants, which in turn results in differences in their associated levels of fitness. Measurable differences in fitness between mutant (evaluated locally) and resident (evaluated over the entire habitat) suggest that it is straightforward to determine when a resident strategy can be invaded by a mutant strategy. We provide detailed derivations of the conditions for such a resident strategy to be locally uninvadable by mutants

playing a similar strategy. A large portion of the discussion is not considered in Broom et al. (2006), and therefore in addition to this the reader is strongly encouraged to consider Scaramangas and Broom (2022) on which much of the presentation of this chapter is based.

Statement and derivation of conditions for local ESS

By construction, in the $\varepsilon \to 0$ limit almost all sites consist entirely of residents playing (r_1, t_1). If the focal individual plays a (local) mutant strategy (r, t) then in a site that contains a cluster of focal relatives we have precisely aN mutants and $(1-a)N$ residents. We arrive at expressions for their associated levels of aversiveness by considering the $\varepsilon \to 0$ limit of the more general expressions involving $\varepsilon > 0$ discussed in the previous section. Indeed, taking the limit as $\varepsilon \to 0$ of I^ε in (2.2.23) and relabelling its limiting value as I (i.e. $I^\varepsilon \to I$ as $\varepsilon \to 0$) we have

$$I = a\frac{N}{n}L(r)H(t) + (1-a)\frac{N}{n}L(r_1)H(t_1)S(|r-r_1|), \tag{2.3.1}$$

which confirms our intuition that there are aN mutants contributing the first term and $(1-a)N$ residents contributing the second term. Similarly, taking the limit as $\varepsilon \to 0$ of (2.2.24) and using the shorthand I_1 to denote its limiting value (i.e. $I_1^\varepsilon \to I_1$ as $\varepsilon \to 0$) we recover the simpler expression for the resident aversiveness

$$I_1 = \frac{N}{n}L(r_1)H(t_1). \tag{2.3.2}$$

The payoffs associated with the mutant and resident are evaluated in much the same way. The mutant payoff is given by

$$P(r, t; r_1, t_1) = \frac{F(t)}{\lambda + D(r)K(t)Q(I)}, \tag{2.3.3}$$

where I is given as in (2.3.1), while the resident payoff is given by

$$P(r_1, t_1; r_1, t_1) = \frac{F(t_1)}{\lambda + D(r_1)K(t_1)Q(I_1)} \tag{2.3.4}$$

with the resident information I_1 given given as in (2.3.2). An important assumption of the model is that functions F, D, K, Q, L and H are assumed to be $\mathcal{C}^l$ with $l \geq 2$ and that the function S is $\mathcal{C}^l$ with $l \geq 2$ sufficiently near the origin (for ESS analysis this is what matters since mutations are taken to be local such that $|r - r_1|$ can be taken to be small). This, together with the fact that I in (2.3.1) depends on quantity $|r - r_1|$ suggests that the mutant fitness $P(r, t; r_1, t_1)$ in (2.3.3) depends is almost everywhere $\mathcal{C}^l$ with $l \geq 2$ except at $r = r_1$ where its (partial) derivative with respect to r is not defined. It should be remarked that while quantity $|r - r_1|$ is not differentiable with respect to r at $r = r_1$ it is continuous at that value so that by extension $P(r, t; r_1, t_1)$ is continuous at $r = r_1$. We remark that while (2.3.1) is evaluated on some local area of the habitat, which contains a cluster of mutants, expression (2.3.2) represents the average level of prey aversiveness as perceived by predators who mostly visit sites consisting entirely of residents. This suggests that while the mutants are present in local areas of the habitat they threaten to invade the resident on the global level. This brings us to the central theme of this section, which is the evolutionary stability of aposematism.

In this manuscript we discuss evolutionarily stability in the spirit in which it had originally been introduced in Maynard Smith and Price, 1973 (see also Maynard Smith, 1982). That is, a local ESS is used to describe a strategy that is resistant to invasion by alternative strategies and in particular, one which when

common could not be invaded by an initially rare alternative strategy.

Definition 2.3.1. *A resident strategy (r_{ESS}, t_{ESS}) is (locally) evolutionarily stable - local ESS - if it is best-response against itself and in particular, if all mutant strategies (r, t) receive lower fitness (P) when interacting with the ESS strategy than does the ESS strategy when interacting with itself. We distinguish the following cases for local ESS*

$[i]$ $(0,0) \in \mathcal{D}_0$ *is a local ESS if* $P(0,0;0,0) > P(r,t;0,0)$

for all $(r,t) \in [0, \delta r] \times [0, \delta t] \setminus (0,0)$;

$[ii]$ $(r_{ESS}, 0) \in \mathcal{D}_1$ *is a local ESS if* $P(r_{ESS}, 0; r_{ESS}, 0) > P(r,t; r_{ESS}, 0)$

for all $(r,t) \in [r_{ESS} - \delta r, r_{ESS} + \delta r] \times [0, \delta t] \setminus (r_{ESS}, 0)$;

$[iii]$ $(0, t_{ESS}) \in \mathcal{D}_2$ *is a local ESS if* $P(0, t_{ESS}; 0, t_{ESS}) > P(r,t; 0, t_{ESS})$

for all $(r,t) \in [0, \delta r] \times [t_{ESS} - \delta t, t_{ESS} + \delta t] \setminus (0, t_{ESS})$;

$[iv]$ $(r_{ESS}, t_{ESS}) \in \mathcal{D}_3$ *is a local ESS if* $P(r_{ESS}, t_{ESS}; r_{ESS}, t_{ESS}) > P(r,t; r_{ESS}, t_{ESS})$

for all $(r,t) \in [r_{ESS} - \delta r, r_{ESS} + \delta r] \times [t_{ESS} - \delta t, t_{ESS} + \delta t] \setminus (r_{ESS}, t_{ESS})$,

where it is assumed that quantities δr and δt are positive and arbitrarily small (i.e. $0 < \delta r \ll 1$ and $0 < \delta t \ll 1$).

Recent progress in evolutionary game theory has given rise to a plethora of terms used to describe a rather limited number of properties relating to evolutionary stability. Among these, we identify resistance to invasion (ESS) and convergence stability as fundamentally separate and refer the reader to Apaloo et al. (2009) and Waxman and Gavrilets (2005) for detailed discussions about the attributes of evolutionary stability and the array of terminologies used to describe these. A strategy is said to be *convergence stable* if it can be approached through a sequence of selectively advantageous mutations and as it happens, resistance to invasion is not sufficient for characterising a local ESS as a likely outcome of evolution by natural selection. Indeed, an ESS strategy might not be favoured if the current population is using a nearby strategy that is not convergence stable.

For cases where there is a continuum of strategies, Eshel and Motro (1981) define the notion of a *continuously stable strategy (CSS)* as an ESS with the additional property that whenever the entire population has a strategy which is close enough to it there will be a selective advantage to some individual strategies which are closer to the CSS. While we do not deal with CSS's in this manuscript, we mention this definition here, as it forms the basis for the distinction between resistance to invasion (ESS) and convergence stability. In fact, a CSS is an ESS that is convergence stable. Subsequent works by Eshel (1983, 1996) and particularly the work by Vincent et al. (1993) have helped incorporate the notion of convergence stability into the fabric of adaptive dynamics, which explores how a population's mean strategy evolves in time through it's *adaptive landscape*. In Abrams et al. (1993) it is showcased that in absence of ESS convergence stability can drive a population to a point of minimum fitness in the adaptive landscape, which prompted subsequent research in adaptive dynamics (see Metz et al., 1995).

Instances in which a mutant strategy does *equally well* against the resident strategy are considered non-

generic and are omitted from Definition 2.3.1 but discussed subsequent to the proof of Theorem 2.3.2. The focus of this manuscript is to provide a static approach to the evolutionary stability of aposematism by considering prey populations that are monomorphic (although see revised definitions in chapters 5 and 6). Such monomorphic populations are considered in the adaptive dynamics framework, in which Definition 2.3.1 would best describe a non-invasible *evolutionarily singular strategy* (ess). We should remark that in the context of adaptive dynamics a non-invasible ess as described in Definition 2.3.1 is related to but not identical to the notion of an ESS, since a strategy that is non-invasible (by neighbouring strategies) may still be invaded by a *distant mutant* - see chapter 13 of Broom and Rychtár (2013). We should also mention that because distant mutants of the type discussed are not considered in this manuscript (mutations are assumed to be local) we can think of uninvasible ess's in this context as identical to local ESSs.

It is immediately clear from Definition 2.3.1 that (in the $\varepsilon \to 0$ limit) if the mutant fitness admits a strict local maximum at the resident value $(r,t) = (r_{ESS}, t_{ESS})$ then (r_{ESS}, t_{ESS}) is a local ESS. It is worth mentioning that if the local maximum is not strict it may include nearby strategies that do equally well against the resident strategy and could be considered non-generic. The reader is directed to chapter 7 of Vincent and Brown (2005) for a more systematic discussion of the *ESS maximum-principle*. Throughout the remainder of this manuscript we rely on a theorem (Theorem 2.3.2) that provides a set of rules for determining whether a resident strategy is a local ESS. We make a number of clarifications about the mutant fitness before proceeding to state and prove this theorem.

The mutant payoff in (2.3.3) is a scalar function of four variables since it depends on the strategy played by the mutant and the strategy played by the resident. In addition to this it is important to remark that on account of the absolute value present in the expression for the mutant aversiveness (2.3.1) the mutant payoff is not differentiable with respect to the first mutant trait when $r = r_1$. Indeed, after suitable positive scaling (see exact derivation in (2.3.57)) the r-derivative of the mutant payoff $\partial_r P(r,t;r_1,t_1)$ one obtains

$$\frac{D'(r)}{D(r)} - a\frac{N}{n}L'(r)H(t)\frac{Q'(I)}{Q(I)} - (1-a)I_1\frac{Q'(I)}{Q(I)}S'(|r-r_1|)\Big(- \mathbb{1}_{(-\infty,r_1)} + \mathbb{1}_{(r_1,+\infty)}\Big), \tag{2.3.5}$$

which is not defined at $r = r_1$ unless $S'(0) = 0$ - a possibility we exclude (see (2.2.5)). An important consequence is that the conditions for maximising the mutant fitness at the resident value (at ESS) cannot be deduced using the standard linearisation techniques from vector calculus. Notice that the mutant payoff exhibits this issue of differentiability only with respect to the first mutant trait (r) while along the second mutant trait (t) this is not the case particularly because the functional forms in Table 2.1 (except for the mentioned bi-variate similarity $\mathcal{S}(r_i, r_j)$) are assumed to be C^l with $l \geq 2$.

As is clear from Definition 2.3.1 the precise conditions for a local maximum depend on where on the boundary-inclusive, right-upper-hand plane $\{(\rho,\tau) : \rho \geq 0, \tau \geq 0\}$ the resident strategy is evaluated. We distinguish between the origin $\mathcal{D}_0 = \{(0,0)\}$, the boundaries $\mathcal{D}_1 = \{\rho > 0, \tau = 0\}$, $\mathcal{D}_2 = \{\rho = 0, \tau > 0\}$ and the interior $\mathcal{D}_3 = \{\rho > 0, \tau > 0\}$. It is shown in Broom et al. (2006) that the non-aversive subset of the interior $\{\rho > 0, \tau \leq t_c\} \subset \mathcal{D}_3$ does not contain local ESSs and we return to this point in due course. The conditions for local ESS are provided in the theorem below, which is central to the remainder of this manuscript.

For clarity we provide limit definitions for the derivatives that are considered in Theorem 2.3.2 below. Let h be positive and arbitrarily small quantity (i.e. $0 < h \ll 1$). We emphasize that $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}$ describes the partial derivative of the mutant payoff with respect to the mutant trait t evaluated at the

resident value $(r,t)=(r_1,t_1)$ with remaining resident traits r_1 and t_1 held fixed. In particular, we write

$$\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} := \lim_{h\to 0}\frac{P(r_1,t_1+h;r_1,t_1)-P(r_1,t_1;r_1,t_1)}{h} \tag{2.3.6}$$

and define higher order derivatives in a similar way. If the resident value is drawn from the boundary $\mathcal{D}_0 \sqcup \mathcal{D}_1 = \{\rho \geq 0, \tau = 0\}$ mutations in t can only be positive so that quantity $\overrightarrow{\partial}_t P(r,t;r_1,0)|_{r=r_1,t=0}$ describes the rate at which the mutant fitness changes in response to changes in the mutant trait for toxicity. We have

$$\overrightarrow{\partial}_t P(r,t;r_1,0)|_{r=r_1,t=0} = \lim_{h\to 0}\frac{P(r_1,h;r_1,0)-P(r_1,0;r_1,0)}{h}. \tag{2.3.7}$$

It is is clear from (2.3.5) that precisely at $r=r_1$ the partial derivative of the mutant payoff with respect to mutant conspicuousness is not defined. Exactly at $r=r_1$ we use the left and right partial derivatives used above, which are defined through

$$\overleftarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1} := \lim_{h\to 0}\frac{P(r_1-h,t_1;r_1,t_1)-P(r_1,t_1;r_1,t_1)}{h} \tag{2.3.8}$$

and

$$\overrightarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1} := \lim_{h\to 0}\frac{P(r_1+h,t_1;r_1,t_1)-P(r_1,t_1;r_1,t_1)}{h}. \tag{2.3.9}$$

We remark that (2.3.8) describes the rate with which the mutant fitness changes in response to a reduction in the level of conspicuousness. The change in fitness associated with a negative step $-h$ along r is given by $P(r_1-h,t_1;r_1,t_1)-P(r_1,t_1;r_1,t_1)$ - i.e. the value after mutation minus the resident value - and can be approximated (to first-order) by $-h\times\overleftarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1}$. For interpreting simulations in chapter 4 we find that it is often more meaningful to use $-\overleftarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1}$ because this accounts for the sign of the mutation step (see also (4.1.15), (4.1.16) and (4.1.19) for the explicit forms used in the simulation some 40 pages ahead). In the theorem that follows we discuss how evaluating derivatives of the mutant fitness at the resident value can be used for determining whether that strategy is a (local) ESS.

Theorem 2.3.2. *Assume that the mutant fitness $P(r,t;r_1,t_1)$ is given through (2.3.3) such that it is almost everywhere $\mathcal{C}^l$ with $l \geq 2$ in the local vicinity of the resident strategy except at $r = r_1$ where it is not r-differentiable, but is continuous at that value. Then the following conditions hold for determining when a resident strategy is locally evolutionarily stable.*

[*i*] *If for* $(0,0) \in \mathcal{D}_0$

$$\overrightarrow{\partial}_t P(r,t;0,0)|_{r=0,t=0} < 0, \quad \textit{and} \tag{2.3.10}$$

$$\overrightarrow{\partial}_r P(r,t;0,0)|_{r=0,t=0} < 0 \tag{2.3.11}$$

then $(0,0)$ *is a local ESS.*

[*ii*] *If for* $(r_1,0) \in \mathcal{D}_1$

$$\overrightarrow{\partial}_t P(r,t;r_1,0)|_{r=r_1,t=0} < 0, \tag{2.3.12}$$

$$\overleftarrow{\partial}_r P(r,t;r_1,0)|_{r=r_1,t=0} > 0 \ \textit{and} \tag{2.3.13}$$

$$\overrightarrow{\partial}_r P(r,t;r_1,0)|_{r=r_1,t=0} < 0 \tag{2.3.14}$$

then $(r_1,0)$ *is a local ESS and is denoted* $(r_{ESS},0)$.

[*iii*] *If for* $(0,t_1) \in \mathcal{D}_2$

$$\partial_t P(r,t;0,t_1)|_{r=0,t=t_1} = 0, \tag{2.3.15}$$

$$\partial_{tt} P(r,t;0,t_1)|_{r=0,t=t_1} < 0 \ \textit{and} \tag{2.3.16}$$

$$\overrightarrow{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} < 0, \tag{2.3.17}$$

then $(0,t_1)$ *is a local ESS and is denoted* $(0,t_{ESS})$

[*iv*] *If for* $(r_1,t_1) \in \mathcal{D}_3$

$$\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} = 0, \tag{2.3.18}$$

$$\partial_{tt} P(r,t;r_1,t_1)|_{r=r_1,t=t_1} < 0, \tag{2.3.19}$$

$$\overleftarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1} > 0 \ \textit{and} \tag{2.3.20}$$

$$\overrightarrow{\partial}_r P(r,t;r_1,t_1)|_{r=r_1,t=t_1} < 0, \tag{2.3.21}$$

then (r_1,t_1) *is a local ESS and is denoted* (r_{ESS},t_{ESS}).

Proof. We show the theorem above by demonstrating in each of cases [*i*] through [*iv*] that the mentioned conditions lead to a local ESS (in the sense of Definition 2.3.1). More specifically, we show that the marginal difference between the mutant fitness and the resident fitness is strictly negative (although see subsequent discussions about resident strategies that do not satisfy the conditions of the theorem).

We begin by considering case $[i]$. In order to demonstrate that the mutant fitness $P(r,t;0,0)$ is lower than the resident fitness $P(0,0;0,0)$ for any mutant value $(r,t) \in [0,\delta r] \times [0,\delta t] \setminus (0,0)$ we express the mutant traits in terms of spherical coordinates x and ϕ, where we let x denote the size of the mutation and ϕ denote the angle between the mutant strategy and the horizontal $t=0$. Unless $\phi = \pi/2$ - a situation we consider separately - the following transformations hold

$$(r,t) \to (x,\phi): \;\; x = \sqrt{r^2+t^2} \;\text{ and }\; \phi = \arctan\frac{t}{r}, \tag{2.3.22}$$

and

$$(x,\phi) \to (r,t): \;\; r = x\cos\phi \;\text{ and }\; t = x\sin\phi. \tag{2.3.23}$$

We can make use of (2.3.23) in order to express the mutant fitness $\mathcal{P}$ in terms of x and ϕ so that

$$\mathcal{P}: \mathbb{R}^{\geq 0} \times \left[0,\frac{\pi}{2}\right] \to \mathbb{R}^{\geq 0} \;\text{ s.t. }\; \mathcal{P}(x,\phi) := P(r = x\cos\phi, t = x\sin\phi; 0,0). \tag{2.3.24}$$

In order to show that the mutant fitness is lower than the resident fitness it remains for us to show that

$$\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) < 0 \;\text{ for all }\; \phi \in [0,\pi/2], \tag{2.3.25}$$

which we show by considering cases $\phi = 0$, $\phi \in (0,\pi/2)$ and $\phi = \pi/2$ separately.

If $\phi = 0$ it follows that $x = r$ so that mutation is along the r-direction and where

$$\mathcal{P}(x,\phi=0) - \mathcal{P}(0,\phi=0) = P(r,0;0,0) - P(0,0;0,0) \approx r \times \underbrace{\vec{\partial}_r P(r,t;0,0)|_{r=0,t=0}}_{<0} < 0. \tag{2.3.26}$$

If $\phi \in (0,\pi/2)$ the incremental difference between the mutant and the resident fitness reads

$$\begin{aligned}
\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) &\approx x\partial_x \mathcal{P}(x,\phi)|_{x=0,\,\phi\in(0,\pi/2)} \\
&= x\partial_x P(r = x\cos\phi, t = x\sin\phi; 0,0)|_{x=0,\,\phi\in(0,\pi/2)} \\
&= \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;0,0)|_{r=0,t=0}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\vec{\partial}_t P(r,t;0,0)|_{r=0,t=0}}_{<0} < 0.
\end{aligned} \tag{2.3.27}$$

Finally, if $\phi = \pi/2$ mutation is along the t-direction so that

$$\mathcal{P}(x,\phi=\pi/2) - \mathcal{P}(0,\phi=\pi/2) = P(0,t;0,0) - P(0,0;0,0) \approx t \times \underbrace{\vec{\partial}_t P(r,t;0,0)|_{r=0,t=0}}_{<0} < 0. \tag{2.3.28}$$

We conclude that we have shown case $[i]$ by showing that (2.3.25) holds.

Similar to case $[i]$ we show $[ii]$ by showing that

$$\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) < 0 \;\text{ for all }\; \phi \in [0,\pi], \tag{2.3.29}$$

where now the mutant strategy is centred at the resident value $(r_{ESS},0) \in \mathcal{D}_1$ so that unless mutation is solely along t (this is case $\phi = \pi/2$, which we consider separately and without making use of the transformations

below) we have

$$(r,t)\to(x,\phi):\ \ x=\sqrt{(r-r_{ESS})^2+t^2}\ \text{ and }\ \phi=\arctan\frac{t}{r-r_{ESS}} \tag{2.3.30}$$

and

$$(x,\phi)\to(r,t):\ \ r=r_{ESS}+x\cos\phi\ \text{ and }\ t=x\sin\phi. \tag{2.3.31}$$

We show inequality (2.3.29) by showing it holds for (the five) cases $\phi=0$, $\phi\in(0,\pi/2)$, $\phi=\pi/2$, $\phi\in(\pi/2,\pi)$ and $\phi=\pi$ separately.

If $\phi=0$ it follows that $x=r-r_{ESS}>0$ so that the difference between mutant and resident fitness reads

$$\begin{aligned}\mathcal{P}(x,\phi=0)-\mathcal{P}(0,\phi=0)&=P(r,0;r_{ESS},0)-P(r_{ESS},0;r_{ESS},0)\\&\approx\underbrace{(r-r_{ESS})}_{>0}\times\underbrace{\overrightarrow{\partial}_rP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{<0}<0.\end{aligned} \tag{2.3.32}$$

If $\phi\in(0,\pi/2)$ it follows that

$$\begin{aligned}\mathcal{P}(x,\phi)-\mathcal{P}(0,\phi)\approx&x\partial_x\mathcal{P}(x,\phi)|_{x=0,\ \phi\in(0,\pi/2)}\\=&x\partial_xP(r=r_{ESS}+x\cos\phi,t=x\sin\phi;r_{ESS},0)\\=&\underbrace{x\cos\phi}_{>0}\times\underbrace{\overrightarrow{\partial}_rP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{<0}+\underbrace{x\sin\phi}_{>0}\times\underbrace{\overrightarrow{\partial}_tP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{<0}<0.\end{aligned} \tag{2.3.33}$$

If $\phi=\pi/2$ mutation is along the t-direction and we have

$$\begin{aligned}\mathcal{P}(x,\phi=\pi/2)-\mathcal{P}(0,\phi=\pi/2)&=P(r_{ESS},t;r_{ESS},0)-P(r_{ESS},0;r_{ESS},0)\\&\approx t\times\underbrace{\overrightarrow{\partial}_tP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{<0}<0.\end{aligned} \tag{2.3.34}$$

Furthermore, if $\phi\in(\pi/2,\pi)$ we have

$$\begin{aligned}\mathcal{P}(x,\phi)-\mathcal{P}(0,\phi)\approx&x\partial_x\mathcal{P}(x,\phi)|_{x=0,\ \phi\in(\pi/2,\pi)}\\=&x\partial_xP(r=r_{ESS}+x\cos\phi,t=x\sin\phi;r_{ESS},0)\\=&\underbrace{x\cos\phi}_{<0}\times\underbrace{\overleftarrow{\partial}_rP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{>0}+\\&+\underbrace{x\sin\phi}_{>0}\times\underbrace{\overrightarrow{\partial}_tP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{<0}<0.\end{aligned} \tag{2.3.35}$$

Finally, if $\phi=\pi$ it follows that $x=r-r_{ESS}<0$ and therefore that

$$P(r,0;r_{ESS},0)-P(r_{ESS},0;r_{ESS},0)\approx\underbrace{(r-r_{ESS})}_{<0}\times\underbrace{\overleftarrow{\partial}_rP(r,t;r_{ESS},0)|_{r=r_{ESS},t=0}}_{>0}<0 \tag{2.3.36}$$

and this concludes case $[ii]$.

Much like in cases $[i]$ and $[ii]$ we show case $[iii]$ by showing that

$$\mathcal{P}(x,\phi) - \mathcal{P}(x=0,\phi) \text{ for all } \phi \in [0,\pi/2] \sqcup [3\pi/2, 2\pi), \tag{2.3.37}$$

where the transformations now read

$$(r,t) \to (x,\phi): \;\; x = \sqrt{r^2 + (t-t_{ESS})^2} \text{ and } \phi = \arctan\frac{t-t_{ESS}}{r} \tag{2.3.38}$$

and

$$(x,\phi) \to (r,t): \;\; r = x\cos\phi \text{ and } t = t_{ESS} + x\sin\phi. \tag{2.3.39}$$

From (2.3.38) it is clear that the transformation should not be utilised when mutation is along the t-direction so that cases $\phi = \pi/2$ and $\phi = 3\pi/2$ will require separate attention. In order to show (2.3.37) we must show that the inequality holds for (the four) cases $\phi = 0$, $\phi \in (0,\pi/2)$, $\phi = \pi/2, 3\pi/2$ and $\phi \in (3\pi/2, 2\pi)$ individually.
If $\phi = 0$ we have

$$\begin{aligned}\mathcal{P}(x,\phi=0) - \mathcal{P}(0,\phi=0) &= P(r,t_{ESS};0,t_{ESS}) - P(0,t_{ESS};0,t_{ESS}) \\ &\approx r \times \underbrace{\vec{\partial}_r P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{<0} < 0.\end{aligned} \tag{2.3.40}$$

If $\phi \in (0,\pi/2)$ we have

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx& x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(0,\pi/2)} \\ =& x\partial_x P(r = x\cos\phi; t = t_{ESS} + x\sin\phi; 0, t_{ESS}) \\ =& \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{<0} + \\ &+ x\sin\phi \times \underbrace{\partial_t P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{=0} < 0\end{aligned} \tag{2.3.41}$$

If $\phi = \pi/2$ or $\phi = 3\pi/2$ we have that $x = t - t_{ESS}$ so that mutation is along the t-direction and thus

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) =& P(0,t;0,t_{ESS}) - P(0,t_{ESS};0,t_{ESS}) \\ \approx& (t-t_{ESS}) \times \underbrace{\partial_t P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{=0} + \\ &+ \frac{1}{2}\underbrace{(t-t_{ESS})^2}_{>0} \times \underbrace{\partial_{tt} P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{<0} < 0.\end{aligned} \tag{2.3.42}$$

Finally, if $\phi \in (3\pi/2, 2\pi)$ we have

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx & x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(3\pi/2,2\pi)} \\ = & x\partial_x P(r = x\cos\phi, t = t_{ESS} + x\sin\phi; 0, t_{ESS})|_{x=0,\ \phi\in(3\pi/2,2\pi)} \\ = & \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{<0} + x\sin\phi \times \underbrace{\partial_t P(r,t;0,t_{ESS})|_{r=0,t=t_{ESS}}}_{=0} < 0.\end{aligned} \tag{2.3.43}$$

We can now safely conclude that statement $[iii]$ of Theorem 2.3.2 holds true by means of inequality (2.3.37).

As with all previous cases, in case $[iv]$ we show that the mutant fitness is lower than the resident fitness by showing that

$$\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) < 0 \ \text{ for all } \ \phi \in [0, 2\pi), \tag{2.3.44}$$

where now the mutant traits are defined in the local vicinity of the resident trait (r_{ESS}, t_{ESS}) so that transformation to spherical coordinates reads

$$(x,\phi) \to (r,t): \ \ x = \sqrt{(r - r_{ESS})^2 + (t - t_{ESS})^2} \ \text{ and } \ \phi = \arctan\left(\frac{t - t_{ESS}}{r - r_{ESS}}\right) \tag{2.3.45}$$

and

$$(r,t) \to (x,\phi): \ \ r = r_{ESS} + x\cos\phi \ \text{ and } \ t = t_{ESS} + x\sin\phi. \tag{2.3.46}$$

We show (2.3.44) by showing that the inequality holds for (the seven) cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2, 3\pi/2$, $\phi \in (\pi/2, \pi)$, $\phi = \pi$, $\phi \in (\pi, 3\pi/2)$ and $\phi \in (3\pi/2, 2\pi)$ separately.

If $\phi = 0$ mutation is solely along the r-direction so that

$$\begin{aligned}\mathcal{P}(x,\phi=0) - \mathcal{P}(0,\phi=0) &= P(r, t_{ESS}; r_{ESS}, t_{ESS}) - P(r_{ESS}, t_{ESS}; r_{ESS}, t_{ESS}) \\ &\approx \underbrace{(r - r_{ESS})}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;r_{ESS},t_{ESS})|_{r=r_{ESS},t=t_{ESS}}}_{<0} < 0.\end{aligned} \tag{2.3.47}$$

If $\phi \in (0, \pi/2)$ the difference between the mutant and the resident fitness reads

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx & x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(0,\pi/2)} \\ = & x\partial_x P(r = r_{ESS} + x\cos\phi, t = t_{ESS} + x\sin\phi; r_{ESS}, t_{ESS}) \\ = & \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;r_{ESS},t_{ESS})|_{r=r_{ESS},t=t_{ESS}}}_{<0} + \\ & + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P(r,t;r_{ESS},t_{ESS})|_{r=r_{ESS},t=t_{ESS}}}_{=0} < 0.\end{aligned} \tag{2.3.48}$$

If $\phi = \pi/2$ or $\phi = 3\pi/2$ mutation is along the t-direction so that $x = t - t_{ESS}$ and thus

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) =& P(r_{ESS}, t; r_{ESS}, t_{ESS}) - P(r_{ESS}, t_{ESS}; r_{ESS}, t_{ESS}) \\ \approx& (t - t_{ESS}) \times \underbrace{\partial_t P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{=0} \\ &+ \frac{1}{2}(t - t_{ESS})^2 \times \underbrace{\partial_{tt} P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{<0} < 0. \end{aligned} \tag{2.3.49}$$

If $\phi \in (\pi/2, \pi)$ we have

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx& x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(\pi/2,\pi)} \\ =& x\partial_x P(r = r_{ESS} + x\cos\phi, t = t_{ESS} + x\sin\phi; r_{ESS}, t_{ESS})|_{x=0,\ \phi\in(\pi/2,\pi)} \\ =& \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{>0} + \\ &+ \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{=0} < 0. \end{aligned} \tag{2.3.50}$$

If $\phi = \pi$ mutation is along r so that $x = r - r_{ESS} < 0$ and thus

$$\begin{aligned}\mathcal{P}(x,\phi=\pi) - \mathcal{P}(0,\phi=\pi) =& P(r, t_{ESS}; r_{ESS}, t_{ESS}) - P(r_{ESS}, t_{ESS}; r_{ESS}, t_{ESS}) \\ \approx& \underbrace{(r - r_{ESS})}_{<0} \times \underbrace{\overleftarrow{\partial}_r P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{>0} < 0. \end{aligned} \tag{2.3.51}$$

If $\phi \in (\pi, 3\pi/2)$ we have

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx& x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(\pi,3\pi/2)} \\ =& x\partial_x P(r = r_{ESS} + x\cos\phi, t = t_{ESS} + x\sin\phi; r_{ESS}, t_{ESS})|_{x=0,\ \phi\in(\pi,3\pi/2)} \\ =& \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{>0} + \\ &+ \underbrace{x\sin\phi}_{<0} \times \underbrace{\partial_t P(r,t; r_{ESS}, t_{ESS})|_{r=r_{ESS}, t=t_{ESS}}}_{=0} < 0. \end{aligned} \tag{2.3.52}$$

Finally, we show that mutations with $\phi \in (3\pi/2, 2\pi)$ lead to lower fitness since

$$\begin{aligned}\mathcal{P}(x,\phi) - \mathcal{P}(0,\phi) \approx & x\partial_x \mathcal{P}(x,\phi)|_{x=0,\ \phi\in(3\pi/2,2\pi)} \\ =& x\partial_x P(r = r_{ESS} + x\cos\phi, t = t_{ESS} + x\sin\phi; r_{ESS}, t_{ESS})|_{x=0,\ \phi\in(3\pi/2,2\pi)} \\ =& \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r P(r,t;r_{ESS},t_{ESS})|_{r=r_{ESS},t=t_{ESS}}}_{<0} \\ &+ \underbrace{x\sin\phi}_{<0} \times \underbrace{\partial_t P(r,t;r_{ESS},t_{ESS})|_{r=r_{ESS},t=t_{ESS}}}_{=0} < 0. \end{aligned} \tag{2.3.53}$$

We conclude that we have shown case $[iv]$ of Theorem 2.3.2 by means of (2.3.44) and through a breakdown of the (seven) cases therein. This concludes the proof of all cases $[i]$ through $[iv]$ and of Theorem 2.3.2. □

We emphasise that Theorem 2.3.2 outlines a direct procedure for determining when a strategy is a local ESS but is not exhaustive. That is, there are possible local ESSs that would not be discovered through Theorem 2.3.2, as well as a number of cases which we classify as non-generic. All of these we discuss presently and by means of example. Let us consider a resident strategy $(r_1, t_1) \in \mathcal{D}_2 \sqcup \mathcal{D}_3$ that violates conditions (2.3.17) and (2.3.15) (and (2.3.21)/(2.3.18) respectively) along the t-direction. This can be realised in three different ways. It could be that both $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} = 0$ and $\partial_{tt} P(r,t;r_1,t_1)|_{r=r_1,t=t_1} = 0$. This is a case we discount as non-generic and classify as inconclusive, since more information would be required about higher order derivatives of the fitness function along t to draw conclusions regarding invasibility. It could be that $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} = 0$ and $\partial_{tt} P(r,t;r_1,t_1)|_{r=r_1,t=t_1} > 0$, which is not a local ESS, since it is possible for mutants to invade. Lastly, it could be that $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} < 0$ or $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} > 0$, both of which describe non-ESSs, since there is risk of invasion by mutants that are less-defended and better-defended respectively. Strategy $(r_1, 0)$ on the boundary $\mathcal{D}_0 \sqcup \mathcal{D}_1$ with $\vec{\partial}_t P(r,t;r_1,0)|_{r=r_1,t=0} > 0$ would likewise describe a non-ESS as it would risk invasion against mutants that are better-defended.

We can also consider a strategy $(0, t_1)$ on the boundary $\mathcal{D}_0 \sqcup \mathcal{D}_2$ that fails condition (2.3.11)/(2.3.20) along the r-direction. This can be achieved in four ways. It could be that $\vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} = 0$ and $\vec{\partial}_r \circ \vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} < 0$, and this could describe a local ESS but is not accounted for by Theorem 2.3.2. The situation with $\vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} = 0$ and $\vec{\partial}_r \circ \vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} = 0$ describes a scenario we would discount as non-generic and which is inconclusive (further information about the fitness function along r is required to draw conclusions about invasibility). The scenario $\vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} = 0$ with $\vec{\partial}_r \circ \vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} > 0$ describes a non-ESS, since such a strategy risks invasion by mutations. Finally, the situation $\vec{\partial}_r P(r,t;0,t_1)|_{r=0,t=t_1} > 0$ clearly describes a non-ESS since such a strategy risks invasion by mutants that are more conspicuous. A similar breakdown of possibilities could be provided for strategies $(r_1, t_1) \in \mathcal{D}_1 \sqcup \mathcal{D}_3$ along r and in either direction.

We devote the remainder of this section to re-writing the functional forms in (2.3.11) through to (2.3.18) in terms of the remaining functional forms of Table 2.1. We begin by evaluating $\partial_r P(r,t;r_1,t_1)$, $\partial_t P(r,t;r_1,t_1)$ and $\partial_{tt} P(r,t;r_1,t_1)$ in the neighbourhood of (r_1, t_1). From the definition of the payoff in (1.3.12) we have

$$\partial_r P(r,t;r_1,t_1) = -\frac{F(t)K(t)}{\big(\lambda + D(r)K(t)Q(I)\big)^2}\left[D'(r)Q(I) + D(r)Q'(I)\partial_r I\right]. \tag{2.3.54}$$

Introducing the (strictly positive) scaling function

$$C : \mathcal{D} \to \mathbb{R}^{>0} : \quad C(r,t) := \frac{(\lambda + D(r)K(t)Q(I))^2}{F(t)D(r)K(t)Q(I)}, \tag{2.3.55}$$

(2.3.54) reads

$$C(r,t)\partial_r P(t,r;t_1,r_1) = -\frac{D'(r)}{D(r)} - \frac{Q'(I)}{Q(I)}\partial_r I, \tag{2.3.56}$$

where scaling by $C(r,t)$ is sign-preserving for all $(r,t) \in \mathcal{D}$ so that

$$\partial_r P(r,t;r_1,t_1) \geq 0 \Leftrightarrow C(r,t)\partial_r P(r,t;r_1,t_1) \geq 0$$

$$\partial_r P(r,t;r_1,t_1) < 0 \Leftrightarrow C(r,t)\partial_r P(r,t;r_1,t_1) < 0.$$

Now from the aversiveness in (2.3.1) we have that

$$\begin{aligned}\partial_r I &= a\frac{N}{n}L'(r)H(t) + (1-a)\frac{N}{n}L(r_1)H(t_1)S'(|r-r_1|)\partial_r|r-r_1| \\ &= a\frac{N}{n}L'(r)H(t) + (1-a)I_1 S'(|r-r_1|)\left(\mathbb{1}_{(-\infty,r_1)} + \mathbb{1}_{(r_1,+\infty)}\right),\end{aligned} \tag{2.3.57}$$

where $S'(|r-r_1|)$ denotes the derivative of the uni-variate similarity function S with respect to its argument. The RHS of (2.3.56) now reads

$$\frac{D'(r)}{D(r)} - a\frac{N}{n}L'(r)H(t)\frac{Q'(I)}{Q(I)} - (1-a)I_1\frac{Q'(I)}{Q(I)}S'(|r-r_1|)\left(-\mathbb{1}_{(-\infty,r_1)} + \mathbb{1}_{(r_1,+\infty)}\right) \tag{2.3.58}$$

and thus we have

$$C(r,t)\partial_r P(r,t;r_1,t_1) = \begin{cases} -\dfrac{D'(r)}{D(r)} - a\dfrac{N}{n}L'(r)H(t)\dfrac{Q'(I)}{Q(I)} + (1-a)I_1\dfrac{Q'(I)}{Q(I)}S'(|r-r_1|), & r < r_1 \\ -\dfrac{D'(r)}{D(r)} - a\dfrac{N}{n}L'(r)H(t)\dfrac{Q'(I)}{Q(I)} - (1-a)I_1\dfrac{Q'(I)}{Q(I)}S'(|r-r_1|), & r > r_1. \end{cases} \tag{2.3.59}$$

Considering mutations that are local amounts to evaluating the derivative on the LHS of the latter at $(r,t) = (r_1,t_1)$ such that we end up with the following directional derivatives

$$C(r_1,t_1)\overleftarrow{\partial}_r P(r,t;r_1,t_1)\big|_{r=r_1,t=t_1} = -\frac{D'(r_1)}{D(r_1)} - aI_1\frac{L'(r_1)}{L(r_1)}\frac{Q'(I_1)}{Q(I_1)} + (1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0) \tag{2.3.60}$$

and

$$C(r_1,t_1)\overrightarrow{\partial}_r P(r,t;r_1,t_1)\big|_{r=r_1,t=t_1} = -\frac{D'(r_1)}{D(r_1)} - aI_1\frac{L'(r_1)}{L(r_1)}\frac{Q'(I_1)}{Q(I_1)} - (1-a)I_1\frac{Q'(I)}{Q(I)}S'(0). \tag{2.3.61}$$

If the resident plays a cryptic strategy (with $r_1 = 0$) mutation in the signalling trait can only lead to more conspicuous mutant types. Naturally, such a resident is presumed to be stable along the r-direction if the single condition

$$-\frac{D'(0)}{D(0)} - aI_1\frac{L'(0)}{L(0)}\frac{Q'(I_1)}{Q(I_1)} - (1-a)I_1\frac{Q'(I)}{Q(I)}S'(0) < 0 \tag{2.3.62}$$

holds true. Strategies with non-zero signalling component ($r_1 > 0$) can give rise to mutants that are either less or more conspicuous. Naturally, such strategies are described as stable in the r-direction if they can resist invasion against the less conspicuous type, which is guaranteed by

$$-\frac{D'(r_1)}{D(r_1)} - aI_1\frac{L'(r_1)}{L(r_1)}\frac{Q'(I_1)}{Q(I_1)} + (1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0) > 0, \tag{2.3.63}$$

and against the more conspicuous mutant type, which is guaranteed through

$$-\frac{D'(r_1)}{D(r_1)} - aI_1\frac{L'(r_1)}{L(r_1)}\frac{Q'(I_1)}{Q(I_1)} - (1-a)I_1\frac{Q'(I)}{Q(I)}S'(0) < 0. \tag{2.3.64}$$

It remains for us to determine explicit expressions for the t-stability conditions. We begin by differentiating the expression for the mutant payoff in (2.3.3) partially with respect to t

$$\partial_t P(r,t;r_1,t_1) = \frac{F'(t)}{\lambda + D(r)K(t)Q(I)} - \frac{F(t)D(r)K'(t)Q(I)}{(\lambda + D(r)K(t)Q(I))^2} + \frac{F(t)D(r)K(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I. \tag{2.3.65}$$

Noting that I is given as in (2.3.1), we evaluate the above at $(r,t) = (r_1,t_1)$ in terms of the scaling function $C(r,t)$ given in (2.3.55)

$$C(r_1,t_1)\partial_t P(r,t;r_1,t_1)\big|_{r=r_1,t=t_1} = \frac{\lambda + D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)}\frac{F'(t_1)}{F(t_1)} - \frac{K'(t_1)}{K(t_1)} - aI_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1)}{Q(I_1)} \tag{2.3.66}$$

and that strategies with $t_1 = 0$ resist invasion from better defended types if

$$\frac{\lambda + D(r_1)K(0)Q(I_1)}{D(r_1)K(0)Q(I_1)}\frac{F'(0)}{F(0)} - \frac{K'(0)}{K(0)} - aI_1\frac{H'(0)}{H(0)}\frac{Q'(I_1)}{Q(I_1)} < 0. \tag{2.3.67}$$

Strategies with $t_1 > 0$ resist invasion from types that are better defended if in addition to satisfying (2.3.17)/(2.3.21)

$$\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1} = 0 \tag{2.3.68}$$

these also satisfy the convexity condition provided in (2.3.18)/(2.3.15)

$$-\partial_{tt} P(r,t;t_1,r_1)\big|_{r=r_1,t=t_1} > 0. \tag{2.3.69}$$

This is equivalent to

$$-C(r_1,t_1)\partial_t\mathcal{A}(r,t)\big|_{r=r_1,t=t_1} + C(r_1,t_1)\partial_t\mathcal{B}(r,t)\big|_{r=r_1,t=t_1} + C(r_1,t_1)\partial_t\mathcal{C}(r,t)\big|_{r=r_1,t=t_1} > 0, \tag{2.3.70}$$

where we have defined $\mathcal{A}(r,t), \mathcal{B}(r,t)$ and $\mathcal{C}(r,t)$ as the three terms on the RHS of the expression for

$\partial_t P(r,t;r_1,t_1)$ (2.3.65), namely

$$\mathcal{A}(r,t) := \frac{F'(t)}{\lambda + D(r)K(t)Q(I)} \tag{2.3.71}$$

$$\mathcal{B}(r,t) := \frac{F(t)D(r)K'(t)Q(I)}{(\lambda + D(r)K(t)Q(I))^2} \tag{2.3.72}$$

$$\mathcal{C}(r,t) := \frac{F(t)D(r)K(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I. \tag{2.3.73}$$

We proceed to evaluating the t-derivatives of these quantities as follows. We have that

$$\partial_t \mathcal{A}(r,t) = \frac{F''(t)}{\lambda + D(r)K(t)Q(I)} - \frac{F'(t)D(r)K'(t)Q(I)}{(\lambda + D(r)K(t)Q(I))^2} - \frac{F'(t)D(r)K(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I, \tag{2.3.74}$$

which, when scaled and evaluated at $(r,t) = (r_1,t_1)$, with $t_1 \neq t_c$ reads

$$-C(r_1,t_1)\partial_t \mathcal{A}(r_1,t_1) = -\frac{\lambda + D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)}\frac{F''(t_1)}{F(t_1)} + \frac{F'(t_1)}{F(t_1)}\frac{K'(t_1)}{K(t_1)} + aI_1\frac{H'(t_1)}{H(t_1)}\frac{F'(t_1)}{F(t_1)}\frac{Q'(I_1)}{Q(I_1)}. \tag{2.3.75}$$

Similarly for $\mathcal{B}(r,t)$ we have that

$$\begin{aligned}\partial_t \mathcal{B}(r,t) =& \frac{F'(t)D(r)K'(t)Q(I)}{(\lambda + D(r)K(t)Q(I))^2} + \frac{F(t)D(r)K''(t)Q(I)}{(\lambda + D(r)K(t)Q(I))^2} + \frac{F(t)D(r)K'(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I \\ &- 2\frac{F(t)D^2(r)K'^2(t)Q^2(I)}{(\lambda + D(r)K(t)Q(I))^3} - 2\frac{F(t)D^2(r)K(t)K'(t)Q(I)Q'(I)}{(\lambda + D(r)K(t)Q(I))^3}\partial_t I.\end{aligned} \tag{2.3.76}$$

Again, scaling and evaluating at $(r,t) = (r_1,t_1)$ such that $t_1 \neq t_c$ amounts to

$$\begin{aligned}C(r_1,t_1)\partial_t \mathcal{B}(r,t)\big|_{r=r_1,t=t_1} =& \frac{F'(t_1)}{F(t_1)}\frac{K'(t_1)}{K(t_1)} + \frac{K''(t_1)}{K(t_1)} + aI_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1)}{Q(I_1)}\frac{K'(t_1)}{K(t_1)} \\ &- \frac{2D(r_1)K'^2(t_1)Q(I_1)}{K(t_1)(\lambda + D(r)K(t)Q(I))} - 2aI_1\frac{H'(t_1)}{H(t_1)}\frac{D(r_1)K'(t_1)Q'(I_1)}{\lambda + D(r)K(t)Q(I)}.\end{aligned} \tag{2.3.77}$$

Finally, we have that

$$\begin{aligned}\partial_t \mathcal{C}(r,t) =& \frac{F'(t)D(r)K(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I + \frac{F(t)D(r)K'(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_t I + \frac{F(t)D(r)K(t)Q''(I)}{(\lambda + D(r)K(t)Q(I))^2}(\partial_t I)^2 \\ &+ \frac{F(t)D(r)K(t)Q'(I)}{(\lambda + D(r)K(t)Q(I))^2}\partial_{tt} I - 2\frac{F(t)D^2(r)K(t)K'(t)Q(I)Q'(I)}{(\lambda + D(r)K(t)Q(I))^3}\partial_t I \\ &- 2\frac{F(t)D^2(r)K^2(t)Q'^2(I)}{(\lambda + D(r)K(t)Q(I))^3}(\partial_t I)^2,\end{aligned} \tag{2.3.78}$$

which when scaled and evaluated at $(r,t)=(r_1,t_1)$ with $t_1 \neq t_c$, reads

$$\begin{aligned} C(r_1,t_1)\partial_t \mathcal{C}(r,t)\big|_{r=r_1,t=t_1} =& aI_1 \frac{H'(t_1)}{H(t_1)} \frac{F'(t_1)}{F(t_1)} \frac{Q'(I_1)}{Q(I_1)} + aI_1 \frac{H'(t_1)}{H(t_1)} \frac{K'(t_1)}{K(t_1)} \frac{Q'(I_1)}{Q(I_1)} \\ &+ a^2 I_1^2 \bigg(\frac{H'(t_1)}{H(t_1)} \bigg)^2 \frac{Q''(I_1)}{Q(I_1)} + aI_1 \frac{H''(t_1)}{H(t_1)} \frac{Q'(I_1)}{Q(I_1)} \\ &- 2aI_1 \frac{H'(t_1)}{H(t_1)} \frac{D(r_1)K'(t_1)Q'(I_1)}{\lambda + D(r_1)K(t_1)Q(I_1)} \\ &- 2a^2 I_1^2 \bigg(\frac{H'(t_1)}{H(t_1)} \bigg)^2 \frac{D(r_1)K(t_1)Q'^2(I_1)}{Q(I_1)(\lambda + D(r_1)K(t_1)Q(I_1))}. \end{aligned} \tag{2.3.79}$$

Summing the terms in expressions (2.3.75), (2.3.77) and (2.3.79) gives

$$\begin{aligned} &- \frac{\lambda + D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)} \frac{F''(t_1)}{F(t_1)} + \frac{K''(t_1)}{K(t_1)} + 2aI_1 \frac{H'(t_1)}{H(t_1)} \frac{Q'(I_1)}{Q(I_1)} \frac{K'(t_1)}{K(t_1)} \\ &+ aI_1 \frac{H''(t_1)}{H(t_1)} \frac{Q'(I_1)}{Q(I_1)} + a^2 I_1^2 \bigg(\frac{H'(t_1)}{H(t_1)} \bigg)^2 \frac{Q''(I_1)}{Q(I_1)} \\ &+ 2\frac{F'(t_1)}{F(t_1)} \frac{K'(t_1)}{K(t_1)} + 2aI_1 \frac{H'(t_1)}{H(t_1)} \frac{F'(t_1)}{F(t_1)} \frac{Q'(I_1)}{Q(I_1)} - 4aI_1 \frac{H'(t_1)}{H(t_1)} \frac{D(r_1)K'(t_1)Q'(I_1)}{\lambda + D(r_1)K(t_1)Q(I_1)} \\ &- \frac{2D(r_1)K'^2(t_1)Q(I_1)}{K(t_1)(\lambda + D(r_1)K(t_1)Q(I_1)} - 2aI_1^2 \bigg(\frac{H'(t_1)}{H(t_1)} \bigg)^2 \frac{D(r_1)K(t_1)Q'^2(I_1)}{\lambda + D(r_1)K(t_1)Q(I_1)} > 0. \end{aligned} \tag{2.3.80}$$

As per the conditions of Theorem 2.3.2 we require the latter to hold in tandem with (2.3.17)/(2.3.21), which amounts to

$$aI_1 \frac{H'(t_1)}{H(t_1)} = \frac{\lambda + D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)} \frac{F'(t_1)}{F(t_1)} - \frac{Q(I_1)}{Q'(I_1)} \frac{K'(t_1)}{K(t_1)}. \tag{2.3.81}$$

Indeed, (2.3.81) can be used in order to simplify inequality (2.3.80). With appropriate scaling the former reads

$$\begin{aligned} -2a^2 I_1^2 \bigg(\frac{H'(t_1)}{H(t_1)} \bigg)^2 \frac{D(r_1)K(t_1)Q'^2(I_1)}{Q(I_1)(\lambda + D(r_1)K(t_1)Q(I_1))} =& -2\bigg(\frac{F'(t_1)}{F(t_1)} \bigg)^2 \frac{\lambda + D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)} \\ &- 2\frac{D(r_1)K'^2(t_1)Q(I_1)}{K(t_1)(\lambda + D(r_1)K(t_1)Q(I_1))} \\ &+ 4\frac{F'(t_1)}{F(t_1)} \frac{K'(t_1)}{K(t_1)}. \end{aligned} \tag{2.3.82}$$

Furthermore, it holds that

$$-4aI_1 \frac{H'(t_1)}{H(t_1)} \frac{D(r_1)K'(t_1)Q'(I_1)}{\lambda + D(r_1)K(t_1)Q(I_1)} = -4\frac{F'(t_1)}{F(t_1)} \frac{K'(t_1)}{K(t_1)} + 4\frac{D(r_1)K'^2(t_1)Q(I_1)}{K(t_1)(\lambda + D(r_1)K(t_1)Q(I_1))}, \tag{2.3.83}$$

and also that

$$2aI_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1)}{Q(I_1)}\frac{F'(t_1)}{F(t_1)}=2\frac{\lambda+D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)}\left(\frac{F'(t_1)}{F(t_1)}\right)^2-2\frac{F'(t_1)}{F(t_1)}\frac{K'(t_1)}{K(t_1)}. \tag{2.3.84}$$

Substituting (2.3.82), (2.3.83) and (2.3.84) into (2.3.80), simplifies the latter, so that (2.3.69) amounts to

$$-\frac{\lambda+D(r_1)K(t_1)Q(I_1)}{D(r_1)K(t_1)Q(I_1)}\frac{F''(t_1)}{F(t_1)}+\frac{K''(t_1)}{K(t_1)}+2aI_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1)}{Q(I_1)}\frac{K'(t_1)}{K(t_1)}$$

$$+aI_1\frac{H''(t_1)}{H(t_1)}\frac{Q'(I_1)}{Q(I_1)}+a^2I_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\frac{Q''(I_1)}{Q(I_1)}>0. \tag{2.3.85}$$

We are now in position to provide explicit conditions for an arbitrary strategy (r_1,t_1) to be local ESS. In keeping with our presentation thus far, we identify four regions and re-express (2.3.11) through to (2.3.18) explicitly.

From Theorem 2.3.2 it follows that if $(0,0)\in\mathcal{D}_0$ satisfies (2.3.11) and (2.3.10) then it is a local ESS. Considering the mutant fitness is given as in (2.3.3) these conditions amount to

$$\frac{\lambda}{D(0)K(0)Q(I_1)}\frac{F'(0)}{F(0)}+\frac{F'(0)}{F(0)}-\frac{K'(0)}{K(0)}-aI_1\frac{H'(0)}{H(0)}\frac{Q'(I_1)}{Q(I_1)}<0,\quad\text{and} \tag{2.3.86}$$

$$-\frac{D'(0)}{D(0)}-aI_1\frac{D'(0)}{D(0)}\frac{Q'(I_1)}{Q(I_1)}-(1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0)<0. \tag{2.3.87}$$

In the proof of case $[i]$ of Theorem 2.3.2 it was shown that condition (2.3.86) guarantees that a resident playing $(0,0)$ can resist invasion holds such a strategy is it was shown that if only condition It should also be mentioned that if only (2.3.86) is satisfied it In Theorem 2.3.2 it was also shown that if a strategy with non-zero signalling component and zero component for defence - i.e. a point of the form $(r_1,0)\in\mathcal{D}_1$ satisfies (2.3.13), (2.3.14) and (2.3.12) then it is a local ESS and we denote it by $(r_{ESS},0)$. On account of (2.3.3) these amount to

$$\frac{\lambda}{D(r_{ESS})K(0)Q(I_1)}\frac{F'(0)}{F(0)}+\frac{F'(0)}{F(0)}-\frac{K'(0)}{K(0)}-aI_1\frac{H'(0)}{H(0)}\frac{Q'(I_1)}{Q(I_1)}<0, \tag{2.3.88}$$

$$-\frac{D'(r_{ESS})}{D(r_{ESS})}-aI_1\frac{D'(r_{ESS})}{D(r_{ESS})}\frac{Q'(I_1)}{Q(I_1)}+(1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0)>0\quad\text{and} \tag{2.3.89}$$

$$-\frac{D'(r_{ESS})}{D(r_{ESS})}-aI_1\frac{D'(r_{ESS})}{D(r_{ESS})}\frac{Q'(I_1)}{Q(I_1)}-(1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0)<0. \tag{2.3.90}$$

From Theorem 2.3.2 it follows that strategies of the form $(0, t_1) \in \mathcal{D}_2$ satisfying (2.3.16), (2.3.17) and (2.3.15) are locally evolutionarily stable and denoted $(0, t_{ESS})$. With the mutant fitness given through (2.3.3) we have shown that these conditions amount to

$$\frac{\lambda}{D(0)K(t_{ESS})Q(I_1)}\frac{F'(t_{ESS})}{F(t_{ESS})} + \frac{F'(t_{ESS})}{F(t_{ESS})} - \frac{K'(t_{ESS})}{K(t_{ESS})} - aI_1\frac{H'(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)} = 0, \tag{2.3.91}$$

$$-\frac{\lambda + D(0)K(t_{ESS})Q(I_1)}{D(0)K(t_{ESS})Q(I_1)}\frac{F''(t_{ESS})}{F(t_{ESS})} + \frac{K''(t_{ESS})}{K(t_{ESS})} + 2aI_1\frac{H'(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)}\frac{K'(t_{ESS})}{K(t_{ESS})}$$

$$+aI_1\frac{H''(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)} + a^2I_1^2\left(\frac{H'(t_{ESS})}{H(t_{ESS})}\right)^2\frac{Q''(I_1)}{Q(I_1)} > 0 \quad \text{and} \tag{2.3.92}$$

$$-\frac{D'(0)}{D(0)} - aI_1\frac{D'(0)}{D(0)}\frac{Q'(I_1)}{Q(I_1)} - (1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0) < 0. \tag{2.3.93}$$

Finally, Theorem 2.3.2 implies that strategies of the form $(r_1, t_1) \in \mathcal{D}_3$ that satisfy (2.3.19), (2.3.20), (2.3.21) and (2.3.18) are locally evolutionarily stable and we denote these as (r_{ESS}, t_{ESS}). We have established that with the mutant fitness given as in (2.3.3) these amount to

$$\frac{\lambda}{D(r_{ESS})K(t_{ESS})Q(I_1)}\frac{F'(t_{ESS})}{F(t_{ESS})} + \frac{F'(t_{ESS})}{F(t_{ESS})} - \frac{K'(t_{ESS})}{K(t_{ESS})} - aI_1\frac{H'(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)} = 0, \tag{2.3.94}$$

$$-\frac{\lambda + D(r_{ESS})K(t_{ESS})Q(I_1)}{D(r_{ESS})K(t_{ESS})Q(I_1)}\frac{F''(t_{ESS})}{F(t_{ESS})} + \frac{K''(t_{ESS})}{K(t_{ESS})} + 2aI_1\frac{H'(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)}\frac{K'(t_{ESS})}{K(t_{ESS})}$$

$$+aI_1\frac{H''(t_{ESS})}{H(t_{ESS})}\frac{Q'(I_1)}{Q(I_1)} + a^2I_1^2\left(\frac{H'(t_{ESS})}{H(t_{ESS})}\right)^2\frac{Q''(I_1)}{Q(I_1)} > 0, \tag{2.3.95}$$

$$-\frac{D'(r_{ESS})}{D(r_{ESS})} - aI_1\frac{D'(r_{ESS})}{D(r_{ESS})}\frac{Q'(I_1)}{Q(I_1)} + (1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0) > 0, \tag{2.3.96}$$

$$-\frac{D'(r_{ESS})}{D(r_{ESS})} - aI_1\frac{D'(r_{ESS})}{D(r_{ESS})}\frac{Q'(I_1)}{Q(I_1)} - (1-a)I_1\frac{Q'(I_1)}{Q(I_1)}S'(0) < 0. \tag{2.3.97}$$

Notice that in the special case that the resident toxicity is equal to the critical level (i.e. $t_1 = t_c$) we would re-write the term $aI_1H'(t_1)/H(t_1)$ as $aN/nL(r_1)H'(t_1)$. Now that we have effectively catalogued the conditions for local ESS (by region) we proceed to remarking the physical significance of these inequalities. This will provide us with physical intuition into the details of the model that is perhaps lacking. Broom et al. (2006) is in fact very in depth and covers a large breadth of topics; perhaps the reason that these details are omitted are to maintain succinctness.

2.4 A marginal fitness interpretation of evolutionary stability

The purpose of this section is two-fold: To discuss the physical significance of the conditions for evolutionary stability and to establish how that discussion leads us into a known result of Broom et al. (2006), namely that strategies with non-zero signalling component that are non-aversive are not evolutionarily stable. In doing so, we consider the conditions along the t-direction and along the r-direction separately. We should

also emphasise that while the content of this chapter is based on Broom et al. (2006) there are a number of differences between this and the latter (indeed the reader is strongly encouraged to read Broom et al., 2006 together with Scaramangas and Broom, 2022 for a more complete view of the model). One key difference is in the interpretation of the habitat as a structure consisting of sites that are territorially-divided among the predators. This means that the perception of predators is shaped by their experience of the prey that reside in the site that they visit so that the presence of prey with marginally varying traits naturally leads to marginal differences in their fitness. The latter also agrees with our perception of the physical world; it guarantees that organisms playing slightly different strategies also have very similar values for the different consequences of their strategies. Finally, much of the discussion found here forms the basis for interpreting the results from the simulations in Chapter 4.

Marginal differences in fitness along the t-direction

We consider mutations in t and explore the landscape of the mutant payoff function in the vicinity of the resident value along this direction. To that end, we consider[8] some mutant strategy with $r = r_1$ and $t \in [t_1 - \delta t, t_1 + \delta t] \setminus \{t_1\}$. To first order, the rate at which the mutant fitness varies with respect to the resident fitness when the mutant trait for defence varies incrementally in the vicinity of the resident value is provided by the (partial) derivative of the mutant payoff along t, whose normalised form is provided in (2.3.66). The terms on the RHS of (2.3.66) correspond to incremental differences associated with the fecundity (first term) and rate of predation of the mutant compared with the resident. The latter are seen as differences in the rate of capture (second term) and in the rate that detected mutants are attacked (third term).

If mutation lead to an increase increase in the defensive trait such that $t \in [t_1, t_1 + \delta t] \setminus \{t_1\}$. Secondary defences are assumed to be costly to the prey that acquire them, which is why better-defended mutants suffer reduced fecundity (sign of first term is negative since $F'(\tau)/F(\tau) < 0$ for all $\tau \geq 0$). On the other hand, better-defended prey are generally harder to capture, which is why the second term has positive sign ($-K'(\tau)/K(\tau) > 0$ for all $\tau \geq 0$). The habitat is territorially divided among predators so that their perception of prey is shaped by their experience of the individuals inhabiting the site that they visit and most groups of predators experience prey playing the resident strategy, since we are working in the $\varepsilon \to 0$ regime. Collectively predators are assumed to have complete experience of prey (this is described as an *equilibrium* for states of hunger, experience etc. in section 2.1) so that incrementally better-defended mutants present in a cluster are perceived as marginally more aversive by the predators visiting their site and are therefore less likely to be attacked. This is accounted for by the third term, which is positive ($Q'(\mathcal{I})/Q(\mathcal{I}) < 0$ for all $\mathcal{I} \in \mathbb{R}$) and represents the marginal fitness advantage associated with reduced rate of attack.

A resident that is completely undefended (i.e. $t_1 = 0$) can give rise to a mutant that is better defended and risks invasion only from such a mutant. We therefore say that such a strategy can resist invasion along the t direction if mutants that are incrementally better-defended exhibit marginally lower fitness compared with the resident. This is precisely when the RHS of (2.3.66) is negative, which holds if the disadvantage associated with reduced fecundity is not outweighed by the reduction in predation. So that, despite its longer life-cycle the more toxic type cannot replace the resident because it still produces fewer offspring within its life-cycle.

For a defended resident (i.e. $t_1 > 0$) it is not sufficient to regard only the first order marginal differences

[8]Implicit in this treatment is that $(r_1, t_1) \in \mathcal{D}_2 \sqcup \mathcal{D}_3$. A discussion involving mutation in t for residents that are undefended with $(r_1, t_1) \in \mathcal{D}_0 \sqcup \mathcal{D}_1$ would be very similar bu in that case we would set $t \in [t_1 - \delta t, t_1] \setminus \{t_1\}$.

in mutant fitness e.g. through (2.3.66). If the RHS of (2.3.66) is negative it is true that such a resident can resist invasion from a better defended mutant but it is also true that is invaded by the less defended mutant. Indeed, if reduced fecundity cannot outweigh reduced predation in the more toxic mutants then form the the differentiability[9] property of the functional forms - assumed to be C^l with $l \geq 2$) for the less toxic type it follows that increased fecundity can outweigh reduced predation and can therefore invade the resident. Likewise, if the RHS of (2.3.66) is positive such a resident can resist invasion from mutant types that are less toxic but is invaded by the more toxic types.

Instead, to guarantee that the defended resident is stable against invasion both from the less and the better defended mutant we impose that the RHS of (2.3.66) is zero and that the second-order marginal differences in fitness, which when solved in tandem with (2.3.94)/(2.3.91) are provided by (2.3.92)/(2.3.95)

Imposing that the RHS of (2.3.92)/(2.3.95) is negative is equivalent to imposing that the landscape of the mutant payoff along t in the vicinity of the resident value is flat (first order) and concave-down (second order) and therefore a local maximum guarantees that the resident can resist invasion from both the less defended and the better defended mutant. So that while to first order we require differences in fitness associated with fecundity to be exactly outweighed by those due to predation to second order it is possible for fecundity disadvantages to outweigh the benefits of reduced predation for $t > t_1$ and for increased fecundity to be outweighed by reduced increased predation for $t < t_1$.

Through this discussion we establish that if strategy $(r_{ESS}, 0)$ is evolutionarily stable then if the prey population played resident strategy with $t_1 > 0$ we expect successive generations of such a population to be invaded by less-defended types (who exhibit marginally higher fitness on account of their fecundity) until the population converges to $t_{ESS} = 0$. If strategy (r_{ESS}, t_{ESS}) is evolutionarily stable with $t_{ESS} > 0$ then (forgetting mutation along r) a population playing $t_1 < t_{ESS}$ is invaded by the better-defended types (who benefit from reduced predation) until after many successive generations the population stabilises to the evolutionarily stable level of defence t_{ESS} from below. Likewise if the prey population played $t_1 > t_{ESS}$ we would expect that they are invaded by less-defended types (who pose a fitness advantage on account of increased fecundity) such that after several successive generations the population converges to the ESS level from above.

Marginal differences in fitness along the r-direction

Let us assume[10] that the residents play some strategy (r_1, t_1) and that mutation occurs in the signalling trait, so that $r \in [r_1 - \delta r, r_1 + \delta r] \setminus \{r_1\}$ and $t = t_1$. This would correspond to a prey population made up almost exclusively of prey playing (r_1, t_1) except for a small number of sites in which there are clusters of mutants who are equally aversive, but whose signalling trait varies marginally with respect to the resident value. Since we have assumed that the functional forms F, D, K, Q, L, H and S (sufficiently near the origin) are $\mathcal{C}^l$ with $l \geq 2$ of Table 2.1 except for $\mathcal{S}$ at $r = r_1$ on account of (2.2.5) are of class C^l with $l \geq 2$ it is safe to conclude that marginal differences in the mutant trait (compared with the resident value) correspond to marginal differences in the mutant fitness.

Remark from definition (2.3.3) of the mutant fitness that the term $D(r)K(t)Q(I)$ represents the predator-induced mortality rate. This suggests that mutants with varying signalling traits experience different fitness because they are predated at different rates compared with their resident counterparts. It follows that a

[9]For instance, if a differentiable function is increasing at a certain point it is equivalent to say that its value increases as the argument increases away from that point or that its value decreases as the argument decreases away from that same point.

[10]Implicit in this discussion is that the resident is conspicuous so that $(r_1, t_1) \in \mathcal{D}_1 \sqcup \mathcal{D}_3$. A discussion involving cryptic residents with $(r_1, t_1) \in \mathcal{D}_0 \sqcup \mathcal{D}_2$ would be simpler and would involve only positive mutation in the signalling trait i.e. $r \in (0, r]$.

strategy is uninvadable in this direction if more/less conspicuous mutants suffer higher rates of predation. It may appear conflicting that a single change in conspicuousness could act to both lower or raise threat due to predation (and therefore fitness). We can gain a deeper understanding of how this is possible by more carefully reviewing the terms involved in the normalised derivatives of the mutant payoff along r provided in (2.3.60) and (2.3.61). From the first term on the RHS of (2.3.60)/(2.3.61) we deduce that the less/more conspicuous type is less/more detectable and thus experiences lower/greater predation on account of this. Our conclusions about the second and third terms are reversed depending on whether $t_1 < t_c$ or $t_1 > t_c$. If $t_1 < t_c$ the second terms in (2.3.60)/(2.3.61) are negative (since $L'(\rho)/L(\rho) > 0$) and the third terms are positive ($S'(x) < 0$ for all $x \geq 0$). That is, the less/more conspicuous type could on the one hand experience lower/higher rates of attack on account being less/more easily recollectable as a type that is attractive and on the other hand a reduced attack rate from not perfectly resembling the majority of attractive residents. If $t_1 > t_c$ the signs of the second and third terms are reversed so that less/more conspicuous types are expected to be attacked more/less on account of being less/more easily recollectable as types that are aversive and both types to be attacked more on account of not perfectly resembling the majority of aversive residents.

Chapter 3

An exploration into the ESS continuum

In the previous chapter we described aposematism as the conspicuous signalling of a defence for the deterrence of predators and assigned two continuous traits $(r_1, t_1) \in \mathbb{R}^{\geq 0} \times \mathbb{R}^{\geq 0}$ to describe the aposematic strategies in prey. This description is most suitable to the mathematical modelling of chemically-defended prey, who make use of external signalling cues in the form of bright skin pigmentation to signal their unprofitability. Although it was discussed that this description most closely describes poison frog species of the *Dendrobatidae* family it is by no means restricted to these. Indeed, the mathematical presentation found herein is plausible and transferable to other forms of secondary defence; it is also natural: the strength of aposematic traits (signalling and defence) are realised as continuously quantifiable quantities that can be varied independently from one another and which together define a two-dimensional strategy space within which the aposematic behaviour of any one organism can be represented by a single point. It is also worth remarking that while the central game of this thesis is non-linear and perhaps more multifaceted than the War of Attrition described in the first chapter, evolutionarily stable outcomes are manifest as a continuum in the strategy space and this is a similarity worth remarking.

The focus of this chapter is to release the model of Broom et al. (2006) from two restricting conjectures that were made upon its publication: (i) that evolutionarily stable levels of signalling and defence define a continuum on which increasing levels of conspicuousness are associated with higher levels of defence; (ii) that to a given level of signalling strength is assigned a unique evolutionarily stable level of defence. As we discuss, empirical and model-based studies are conflicting regarding how aposematic traits are related to one another in nature. Although it is true that the majority of works allude to a positive relationship between these quantities (i.e. more conspicuous animals are better defended), this is by no means a definitive conclusion. Even if this were the case, there is no one accepted mechanism by which this is facilitated (although we do discuss a number of these). In this section we demonstrate that both positively and negatively correlated combinations of traits can achieve evolutionarily stable outcomes and further, that for a given level of signal strength there can be more than one optimal level of defence. As we discover, our findings could bare interest both to the mathematical-oriented readership and to the biologist alike. It is especially intriguing that the model of Broom et al. (2006) can account for such a vast range of outcomes.

3.1 The monotonicity of the ESS continuum

We spent a considerable portion in the closing section of the previous chapter demonstrating how and why the conspicuous signalling of non-aversive levels of defence i.e. strategies drawn from $\{\rho > 0, \tau \leq t_c\} \subset \mathcal{D}_2 \sqcup \mathcal{D}_3$ is not evolutionarily stable. To that, much of the discussion of this chapter is focused on the conspicuous signalling of aversive strategies and the possibility of such outcomes co-existing (as ESSs) alongside crypsis. We recall from Theorem 2.3.2 that for a strategy $(r_1, t_1) \in \mathcal{D}_2 \sqcup \mathcal{D}_3$ with non-zero levels of defence to be locally evolutionarily stable it must satisfy equality (2.3.17)/(2.3.21).

We emphasize that throughout this manuscript mutations are assumed to be local so that mutant traits are defined in the local vicinity of the resident value - see (2.2.17) through to (2.2.20) for a careful definition of the term local vicinity. Since there are two independent choices for the mutant traits (providing these are drawn locally) and two independent choices for the resident traits such that all four choices are determining for the value of mutant fitness it follows that this quantity, which is given in (2.3.3) is a scalar-valued function of four variables. Since the conditions of Theorem 2.3.2 involve derivatives of the mutant fitness evaluated at the resident strategy it follows that (2.3.17)/(2.3.21) is an equality involving the resident traits r_1 and t_1 only (non-dependence on mutant traits is also immediately clear from the limit definition provided in (2.3.6)) and hence specifies a curve in $\mathcal{D}_2 \sqcup \mathcal{D}_3$.

In section 2.4 it was explained that resident strategies satisfying (2.3.17)/(2.3.21) exhibit the unique property that mutations in the trait for defence lead to an exact trade-off (evaluated to first order) between reproduction and life-span, which can be interpreted as a form of equilibrium on the level of the trade-offs. That is to say, better-defended mutants produce fewer offspring but survive more attacks such that on average the number of offspring they produce per life cycle (i.e. their fitness) is the same as that of their resident counterparts. Strategies satisfying (2.3.17)/(2.3.21) - denoted (r_1^*, t_1^*) - are elements of the curve

$$\left\{(r_1^*, t_1^*) : \partial_t P(r, t; r_1, t_1)|_{r=r_1^*, t=t_1^*} = 0\right\}, \tag{3.1.1}$$

on which the equilibrium level of defence is defined implicitly in terms of the conspicuousness. Evolutionarily stable outcomes are manifest as sections of this continuum specified through the remaining inequalities in (2.3.16) through to (2.3.15) and (2.3.19) through to (2.3.18).

Since we deal exclusively with trait mutations that are local, the landscape of the mutant fitness in the immediate vicinity of the resident value is sufficient to determine the evolutionary stability of the associated resident strategy - see Definition 2.3.1. Mutations are continuous and taken to be so small that information about the derivatives of the mutant payoff at the resident value can be used to determine (through the use of Theorem 2.3.2) whether the associated resident strategy is an ESS. As the conditions in Theorem 2.3.2 involve taking derivatives at the resident value only, we make use of the following shorthand notation

$$\partial_t P(r_1, t_1) \leftrightarrow \partial_t P(r, t; r_1, t_1)|_{r=r_1, t=t_1}, \tag{3.1.2}$$

where the limit definition of the RHS of (3.1.2) is given in (2.3.6). The choice of notation in (3.1.2) suppresses the resident traits, which is sensible (considering the derivatives of the mutant fitness are evaluated at the resident value) but raises a number of important technical questions, which we discuss.

The functions F, D, K, Q, L, H and S are all $\mathcal{C}^l$ with $l \geq 2$ sufficiently near the origin. From condition (2.2.5) it follows that $\mathcal{S}(r, r_1) = S(|r - r_1|)$ and in particular that (with respect to the mutant traits) the mutant fitness is almost everywhere $\mathcal{C}^l$ with $l \geq 2$, except for $r = r_1$. An immediate consequence of the latter

is that (for fixed levels of r_1 and t_1) the quantity $\partial_t P(r,t;r_1,t_1)$ is $\mathcal{C}^{l-1}$ with $l \geq 2$ almost everywhere (except at $r = r_1$). The latter, together with the fact that $|r - r_1|$ is continuous at $r = r_1$ implies that $\partial_t P(r,t;r_1,t_1)$ is continuous at the resident value, suggesting that $\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}$ corresponds to $\partial_t P(r,t;r_1,t_1)$ evaluated at $(r,t) = (r_1,t_1)$ and in particular that the shorthand notation in (3.1.2) is well-defined.

Thus far, we have discussed the properties of the mutant fitness $P(r,t;r_1,t_1)$ by assuming that the resident traits are fixed and allowing the mutant traits to vary in the local vicinity of the resident value. We should stress that once the quantity $\partial_t P(r,t;r_1,t_1)$ has been evaluated at the resident value the resulting quantity $\partial_t P(r_1,t_1)$ is a function of the resident traits only.

Observing that $\partial_t P(r_1,t_1)$ is composed of $\mathcal{C}^{l-1}$ functions, where S and $|r - r_1|$ do not feature, it follows that this is a $\mathcal{C}^{l-1}$ function (with respect to the resident traits). In particular, this suggests that the curve (3.1.1) is given as the zero level set of a bi-variate $\mathcal{C}^{l-1}$ function involving only the resident traits r_1 and t_1. Zero-level sets are invariant under non-zero scaling, so that setting

$$g_1(r_1,t_1) := C(r_1,t_1)\partial_t P(r_1,t_1), \tag{3.1.3}$$

with non-zero scaling factor

$$C(r_1,t_1) := \frac{(\lambda + D(r_1)K(t_1)Q(I_1))^2}{F(t_1)D(r_1)K(t_1)Q(I_1)} \tag{3.1.4}$$

suggests that

$$g_1(r_1^*,t_1^*) = 0 \Leftrightarrow \partial_t P(r_1^*,t_1^*) = 0. \tag{3.1.5}$$

From the latter we conclude that the equilibrium curve (3.1.1) can be expressed as

$$\{(r_1^*,t_1^*) : g_1(r_1^*,t_1^*) = 0\}. \tag{3.1.6}$$

The scaling parameter $C(r_1,t_1)$ is composed exclusively of the functions F, D, K, Q, H and L and is therefore $\mathcal{C}^l$ with $l \geq 2$. Since $\partial_t P(r_1,t_1)$ is $\mathcal{C}^{l-1}$ it follows that the product $C(r_1,t_1)\partial_t P(r_1,t_1)$ i.e. $g_1(r_1,t_1)$ is also everywhere $\mathcal{C}^{l-1}$ with $l \geq 2$. The scaling parameter is positive so that additional restrictions on curve (3.1.1) for local ESS - see (2.3.92) through to (2.3.93) as well as (2.3.95), (2.3.96) and (2.3.97) - are maintained. We have thus concluded that local ESSs are manifest as subsets of the curve (3.1.6), which is given as the zero level-set of a (bi-variate) $\mathcal{C}^{l-1}$ function with $l \geq 2$ involving the resident traits. We build on this observation in the discussion that follows.

While it is not generally possible to arrive at an explicit expression for the equilibrium level of defence in terms of the conspicuousness in (3.1.6), this can be achieved (locally) providing the mentioned curve is given through the zero-level set of a continuously differentiable function (*Implicit Function Theorem*). The Implicit Function Theorem is an existence theorem and in two dimensions states precisely that a curve in $\mathbb{R}^2$ defined implicitly as the zero-level set of some continuously differentiable function - see (3.1.7) - can be expressed locally as the graph of some continuously differentiable function. Furthermore, the theorem provides such that the slope of the line tangent to this curve is calculated through the (3.1.9). Given below is the theorem as provided in Baxandall and Liebeck (1986) for two dimensions.

Theorem 3.1.1. (Implicit Function Theorem in $\mathbb{R}^2$) *Let $F : D \subseteq \mathbb{R}^2 \to \mathbb{R}$ be a real-valued, continuously*

differentiable function defined in a neighbourhood D of a point $(a,b) \in \mathbb{R}^2$, such that

$$F(a,b) = 0, \tag{3.1.7}$$

$$\partial_y F(a,b) \neq 0. \tag{3.1.8}$$

Then there exists a neighbourhood N of $a \in \mathbb{R}$, a neighbourhood M of $b \in \mathbb{R}$ and a continuously differentiable function $f : N \subseteq \mathbb{R} \to \mathbb{R}$, such that

$$f(a) = b \quad \textit{and} \quad f(N) \subseteq M; \tag{3.1.9}$$

for each $x \in N$ the equation $F(x,y) = 0$ is uniquely solved by $y = f(x) \in M$, provided that the possible values of y are restricted to lie within M. Moreover, the derivative of f is given by

$$f'(t) = -\frac{\partial_x F(t, f(t))}{\partial_y F(t, f(t))}, \quad t \in N. \tag{3.1.10}$$

Having demonstrated that the equilibrium curve is provided as the zero level set of continuously differentiable function $g_1(r_1, t_1)$ it follows that from (3.1.10) of Theorem 3.1.1 that the slope of the line tangent to point (r_1^*, t_1^*) of the equilibrium curve is provided by

$$-\frac{\partial_{r_1} g_1(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*}}{\partial_{t_1} g_1(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*}}. \tag{3.1.11}$$

Since the zero-level sets of $g_1(r_1,t_1)$ and $\partial_t P(r_1,t_1)$ are identical we would expect their associated equilibrium curves to be indistinguishable (by definition). Naturally, we also expect the slope of the line tangent to some element (r_1^*, t_1^*) of the level set of $g_1(r_1,t_1)$ to be the same the slope of the line tangent to the same strategy perceived as an element of the zero level set of $P(r_1,t_1)$. Indeed, we have

$$\begin{aligned}
&-\frac{\partial_{r_1} g_1(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*}}{\partial_{t_1} g_1(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*}} \\
&= -\frac{\partial_{r_1} C(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*} \times \partial_t P(r,t;r_1,t_1)|_{r=r_1=r_1^*,t=t_1=t_1^*} + C(r_1^*,t_1^*) \times \partial_{r_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*}}{\partial_{t_1} C(r_1,t_1)|_{r_1=r_1^*,t_1=t_1^*} \times \partial_t P(r,t;r_1,t_1)|_{r=r_1=r_1^*,t=t_1=t_1^*} + C(r_1^*,t_1^*)\partial_{t_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*}} \\
&= -\frac{\partial_{r_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*}}{\partial_{t_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*}},
\end{aligned} \tag{3.1.12}$$

as required, where the last step follows from the fact that

$$\partial_t P(r,t;r_1,t_1)|_{r=r_1=r_1^*,t=t_1=t_1^*} = 0. \tag{3.1.13}$$

For purposes of clarity we remark that

$$\partial_{t_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*} \neq \partial_{t_1 t} P(r,t;r_1,t_1)|_{r=r_1=r_1^*;t=t_1=t_1^*}, \tag{3.1.14}$$

which follows from the fact that

$$\partial_{t_1}\left[\partial_t P(r,t;r_1,t_1)|_{r=r_1,t=t_1}\right]|_{r_1=r_1^*,t_1=t_1^*} = \partial_{t_1 t} P(r,t;r_1,t_1)|_{r=r_1=r_1^*;t=t_1=t_1^*} + \partial_{tt} P(r,t;r_1,t_1)|_{r=r_1=r_1^*;t=t_1=t_1^*}.$$

Resident strategies (r_1, t_1) are chosen from the boundary-inclusive, right-upper-hand plane $\mathcal{D} = \{\rho \geq 0, \tau \geq 0\}$ where the conditions for maximising mutant payoff (over its local vicinity) are different at the origin $\mathcal{D}_0 = \{\rho = 0, \tau = 0\}$ to what these are on the boundaries $\mathcal{D}_1 = \{\rho > 0, \tau = 0\}$, $\mathcal{D}_1 = \{\rho = 0, \tau > 0\}$ or the interior $\{\rho > 0, \tau > 0\}$. It is shown in Broom et al. (2006) that the region $\{\rho > 0, \tau \leq t_c\}$ does not contain local ESSs; presently we focus on aversive strategies with non-zero signalling component drawn from $\{\rho > 0, \tau > t_c\}$. Such strategies are evolutionarily stable if conditions (2.3.94) trough to (2.3.97) hold.

Alongside strategies with non-zero signalling component we also consider evolutionarily stable levels of crypsis. If the curve defined through (2.3.94) intersects the $r_1 = 0$ axis and the associated intersection point satisfies the convexity condition (2.3.92) and (2.3.93) then the associated cryptic solution is a local ESS. As a candidate for crypsis, we also consider the origin, which is evolutionarily stable if it satisfies (2.3.86) and (2.3.87).

Theorem 3.1.1 provides the conditions under which the equilibrium toxicity provided through (3.1.6) can be expressed (locally) as the graph of a function of one variable such that a tangent line (and slope) are well-defined. While we have demonstrated that the equilibrium curve satisfies the conditions required by Theorem 3.1 (provided that the functional forms in Table 2.1 are C^l with $l \geq 2$) one should not be confused and conclude that the curve can be expressed as the graph of a single uni-variate (C^{l-1}) function over the entire domain. To be exact, the theorem only provides the conditions that the slope of the tangent is well-defined and as we discover it is not generally possible to arrive at an explicit relationship between the evolutionarily stable levels of defence and the signalling of that defence (while this is achieved in the example that follows, the relationship is not straightforward). We should also highlight that there are two critical cases to consider. These are: (a) the tangent line is vertical - this is the case if $\partial_{t_1} g_1(r_1, t_1)|_{r=r_1^*, t=t_1^*} = 0$ vanishes and (b) there are multiple curves passing though the given point. Case (b) is prohibited by the restriction that $g_1(r_1, t_1)$ is C^l with $l \geq 2$.

In Broom et al. (2006) it was assumed that the equilibrium curve can be expressed as the graph of some continuously differentiable function and in Broom et al. (2008) examples for the functional forms in Table 2.1 were provided, which through (3.1.15) lead to such a picture. Presently, we are not rejecting that this is a possibility but contend that a more general procedure is required to account for a larger plethora of outcomes. We explore an explicit set of example functions that relies on this discussion.

In this chapter we deal explicitly with the case $\lambda = 0$ in which the equilibrium curve given by $g_1(r_1^*, t_1^*) = 0$ in (3.1.6) amounts to

$$\frac{F'(t_1^*)}{F(t_1^*)} - \frac{K'(t_1^*)}{K(t_1^*)} - aI_1 \frac{Q'(I_1)}{Q(I_1)} \frac{H'(t_1^*)}{H(t_1^*)} = 0, \tag{3.1.15}$$

where the LHS is C^{l-1} with $l \geq 2$.

3.2 An explicit procedure for determining local ESSs

In this section we demonstrate that the model of Broom et al. (2006) can be extricated from two restrictive conjectures that were made upon its publication. In particular, we show that both positive and negative combinations of aposematic traits can achieve locally evolutionarily stable outcomes and further that for given level of conspicuousness there can be more than one optimal level of defence. A number of empirical (and model-based studies) suggest that aposematic traits co-evolve so that increased conspicuousness is coupled with increased unpalatability. Although this is a sensible assumption to make (more conspicuous prey are expected to suffer higher predator attack rates hence necessitating increased levels of defence), we

contend that this is neither indicative of the full breath of real physical systems that have been considered nor is it a necessary prediction of Broom et al. (2006).

To showcase our findings we determine evolutionarily stable outcomes for a set of example functions that are much like those considered in Broom et al. (2008), except for the predator attack probability Q, which now exhibits a rather less-idealised dependence on the perceived aversiveness. The section is structured as follows: we demonstrate the general process for solving the conditions for (local) evolutionary stability and propose a partitioning of the parameter space from which emerge three distinct types of solution. Choices of parameters are made from within each partition and explicit continua of evolutionarily stable solutions are depicted (second subsection) in the strategy space.

Example functions

We begin by assigning examples of functions to the general forms introduced in section 2 as in Broom et al. (2008), but with Q now showing stronger dependence on aversive information

$$F(t) := f_0 \exp(-ft); \quad H(t) := t - t_c; \quad K(t) := \frac{k_0}{1+kt};$$

$$L(r) = D(r); \quad Q(\mathcal{I}) := q_0 \exp\left(-q\mathcal{I}^2\right); \quad S(x) = max(1-vx, 0); \quad \lambda = 0. \tag{3.2.1}$$

While the functional form for $Q(\mathcal{I}) = q_0 \exp(-q\mathcal{I})$, which is utilised in Broom et al. (2006) and later in Broom et al. (2008) is perfectly plausible, it is also quite special in the sense that for this the derived quantity $-Q'/Q \circ (\mathcal{I})$ is equal to the constant value q for all levels of the perceived aversiveness. The quantity $-Q'/Q$ represents the rate at which a marginal change in the perceived aversiveness impacts the attack probability as a proportion of the absolute probability of attack (i.e. the slope of the line tangent to the graph of Q divided by the value of the function at $\mathcal{I}$).

More importantly, the term $-Q'/Q$ which has previously been treated as a constant is ubiquitous in the conditions for evolutionary stability - see (2.3.86) through to (2.3.97) - and has therefore lead to a perhaps simplified family of cases in which the level of defence at equilibrium can be expressed explicitly in terms of the conspicuousness. We are not arguing that a squared-dependence on $\mathcal{I}$ in the exponent of Q is necessarily more plausible than a linear one and in fact we do employ the latter in chapters 4 and 6. Instead, we argue that precisely because with a squared-dependence we have $-Q'/Q \circ (I_1) = qI_1$ - where the RHS depends on both resident traits - this provides a good framework within which to explore a larger plethora of examples in which the relationship between aposematic traits is naturally more complex. The example we consider here is one of (perhaps) many that necessitate the more careful mathematical details provided in the previous section. In addition, while $S(x) = max(1-vx, 0)$ is not C^2 everywhere, it is C^2 sufficiently near the origin, which, as we establish is what matters for ESS analysis.

Interestingly, this modification in Q implies that investment in conspicuousness is more effective at reducing predator attack rate (given by the product $D(r)Q(I)$) when compared with the original setup considered in Broom et al. (2008) and hence provides good grounds within which to challenge the notion that more conspicuous prey are attacked more (and thus need be better defended). Indeed, Darst et al. (2006) studied different genera of the *Dendrobatidae* (poison frog) family and observed that the most conspicuous morphs are the least toxic, while the least conspicuous ones were the most toxic. It was suggested therein that once aposematism has become established in a population of prey that the aposematic traits become decoupled so that arbitrary combinations of these can provide optimal protection against predation and

indeed that potentially costly unprofitability may be traded off in favour of bright colouration so that optimal investment in secondary defence will diminish when more cost-effective conspicuousness evolves as a primary defence. Notably, similar observations were made by I. J. Wang (2011), who looked at different populations of the *Oophaga granulifera* species within the *Dendrobatidae* family and considered the mechanisms proposed by Darst et al. (2006) to justify his findings.

Darst et al. (2006) were the first to uncover negatively correlated aposematic traits in nature and indeed the first to ever provide a sound explanation for this (using a differential costs analysis based around optimising energy expenditure to reduce predator attack rate). We presently recover solutions in which aposematic traits appear decoupled and negatively correlated, with the underlying mechanism detailed in terms of a robust mathematical framework. In particular, we suggest that traits are selected so as to optimise fitness in the sense of the conditions detailed in the previous section, which we presently solve for the functional forms provided here.

Local evolutionary stability

Substitution of (3.2.1) into (3.1.15) provides an explicit expression for (3.1.15) in terms of general rate of detection

$$-f + \frac{k}{1+kt_1^*} + 2\alpha D^2(r_1^*)(t_1^* - t_c) = 0, \tag{3.2.2}$$

where the quantity $\alpha = aqN^2/n^2$ has been introduced. For the remainder of the discussion, changes in α are attributed to changes in the fraction N/n representing the relative proportion of prey to predators, with a and q held fixed.

For (3.2.2) we note two interesting facts: first, for given level of conspicuousness there can be (at most) two solutions for t_1^*. Second, the solutions t_1^* can be both increasing and decreasing with increasing conspicuousness and changes in its monotonicity occur on vertices at which the tangent to the curve is vertical. We can see the first by noticing that with appropriate scaling (3.2.2) amounts to

$$t_1^{*2} + A(r_1^*)t_1^* + B(r_1^*) = 0, \tag{3.2.3}$$

with

$$A(r_1^*) = -\frac{f}{2\alpha D^2(r_1^*)} + \frac{1}{k} - t_c \qquad \text{and} \qquad B(r_1^*) = \frac{1}{2\alpha D^2(r_1^*)}\left[1 - \frac{f}{k}\right] - \frac{t_c}{k}.$$

With remaining parameters specified, fixing r_1, plotting the concave-up parabola given by the LHS of (3.2.3) over $t_1 > t_c$ and repeating the process for all r_1^0 is an intuitive way of visualising the surface on the LHS of (3.2.3). The surface of (3.2.3) therefore consists of a family of concave-up parabolas in t_1 parametrized by r_1 and crosses the (r_1, t_1)-plane along the roots

$$t_1^*(r_1^*) = \frac{-A(r_1^*) \pm \sqrt{A^2(r_1^*) - 4B(r_1^*)}}{2}. \tag{3.2.4}$$

We should remark that while (3.2.4) provides an explicit relationship between the level of signalling and the level of defence being signalled at equilibrium (3.2.2), expression (3.2.4) is less useful from a practical standpoint. In fact, we proceed to drawing a number of interesting conclusions about the form of (3.2.2) through the Implicit Function Theorem. In agreement with the notation in (3.1.6), the equilibrium curve can be identified with the set

$$\{(r_1^*, t_1^*) : r_1^* > 0;\, t_1^* = t_1^*(r_1^*)\},$$

which is generated by plotting the roots of (3.2.3) for over $r_1^* \geq 0$.

For the second fact, we apply (3.1.11) to (3.2.2) - it should be evident that (3.2.2) indeed satisfies the necessary conditions for theorem 3.1.1 - and deduce that the slope of the line tangent to any point (r_1^*, t_1^*) of the curve (3.2.2) is given by

$$\frac{4\alpha D(r_1)D'(r_1^*)(t_1^* - t_c)}{\left(\frac{k}{1+kt_1^*}\right)^2 - 2\alpha D^2(r_1^*)}. \tag{3.2.5}$$

Substitution of (3.2.2) into the denominator of (3.2.5) above yields the equivalent expression

$$\frac{-4\alpha}{f}\frac{D(r_1^*)D'(r_1^*)(t_1^* + 1/k)^2(t_1^* - t_c)^2}{t_1^{*2} + 2t_1^*\left(\frac{1}{k} - \frac{1}{f}\right) + \frac{t_c}{f} + \frac{1}{k^2} - \frac{1}{fk}}. \tag{3.2.6}$$

The numerator of the fraction in (3.2.6) is strictly negative (except at $t_1^* = t_c$ - a level of defence, which is not evolutionarily stable when signalled conspicuously), whereas the sign of the denominator is not restricted in this manner. Indeed, the denominator is a concave-up, second-order polynomial which has two real roots (providing $b < 1$) that are given by

$$T_{SIGN}^{\pm} = \frac{1}{f} - \frac{1}{k} \pm \frac{1}{f}\sqrt{1-b}, \quad \text{where} \quad b := f\left(\frac{1+kt_c}{k}\right). \tag{3.2.7}$$

Clearly if $b = 1$ the polynomial has one root only, while if $b > 1$ the polynomial is strictly positive. This implies that expression (3.2.7) provides a natural partitioning of the parameter space with respect to b (at $b = 1$) so that there are two distinct descriptions: $b < 1$ and $b > 1$ with $b = 1$ admitting a border-line case.

Parameters f, k and t_c with $b \geq 1$ correspond to instances when the polynomial in the denominator of (3.2.6) is positive and where the equilibrium curve in (3.2.2) is manifest as a continuum wherein t_1^* decreases with increasing levels r_1^*. In this case ($b \geq 1$), the strategy space is partitioned so that the t_1^* is decreasing with respect to r_1^* whenever $0 < t_1 < T_{SIGN}^-$ or $t_1 > T_{SIGN}^+$ and increasing when $T_{SIGN}^- < t_1 < T_{SIGN}^+$. Intersections of (3.2.2) with the horizontal lines $T_{SIGN}^{\pm}$ are realised as vertices at which a line tangent to the curve is vertical. This is where the branches of the curve described in (3.2.4) meet, which can either be achieved at an intersection with T_{SIGN}^- where the curve exhibits a local minimum in the r-direction (r_{min}-type vertex) or at an intersection with T_{SIGN}^+, where the curve exhibits a local maximum in the r-direction (r_{max}-type vertex).

Considering (3.2.1) it is clear that for given level of investment in defence increasing values of f are associated with higher reductions in fecundity, while increasing values of k are associated with reduced likelihood of an attack resulting in capture. Therefore b can be interpreted as an honest measure of prey sensitivity to investment in toxicity such that prey individuals with $b < 1$ can be thought of as t-insensitive, while those with $b > 1$ are t-sensitive. Ceteris paribus, when b is high, low fecundity is not compensated by lower predation.

Suitable substitution of f in terms of b, k and t_c allows us to re-write (3.2.7) as

$$T_{SIGN}^{\pm} = \left(t_c + \frac{1}{k}\right)\left(\frac{1 \pm \sqrt{1-b}}{b}\right) - \frac{1}{k}. \tag{3.2.8}$$

This reformulation indicates that T_{SIGN}^- assumes ever-increasing values over the interval $\left(0.5t_c - 0.5/k, t_c\right]$

as b increases, while T^{+}_{SIGN} assumes ever-decreasing values over the interval $[t_c, +\infty)$ as b increases. In particular, this shows that the width of the region of increasing solutions $(T^{-}_{SIGN}, T^{+}_{SIGN})$ is greatest for $b \approx 0$ and least for $b \approx 1$ and shrinks monotonically as b increases in between these end values. Furthermore, since $T^{-}_{SIGN} < t_c$ for all choices of b, the equilibrium curve cannot exhibit r_{min}-type vertices in the aversive region, so that only r_{max}-type vertices can be expected for $b < 1$.

Thus far, we have established an effective partitioning of the parameter space and utilised Theorem 3.1.1 to classify the solutions. However, Theorem 3.1.1 is an existence theorem; in order to determine the set of strategies that satisfy the equilibrium condition (3.2.2) we proceed by the suitable re-arrangement

$$\frac{1}{D^2(r_1^*)} = \frac{2\alpha}{f}\frac{(t_1^* + 1/k)(t_1^* - t_c)}{t_1^* - \frac{1}{f} + \frac{1}{k}}. \tag{3.2.9}$$

The latter is convenient because solutions can be understood as intersections of the r_1-dependent LHS and the t_1-dependent RHS, which can be plotted separately as graphs of uni-variate functions. Indeed, this process is outlined explicitly in Figures 3.3 and 3.4. Using the substitution

$$f = \frac{bk}{1 + kt_c}$$

equality (3.2.9) amounts to

$$\frac{1}{D^2(r_1^*)} = \frac{2\alpha(1 + kt_c)}{bk}\frac{(t_1^* + 1/k)(t_1^* - t_c)}{t_1^* - t_*}. \tag{3.2.10}$$

with

$$t_* := \frac{1}{k}\left(\frac{1}{b} - 1\right) + \frac{t_c}{b}.$$

It is evident from the definition of t_* that we have recovered the same partitioning of the parameter space at $b = 1$. Indeed, if $b < 1$ then $t_* > t_c$ and the denominator is zeroed at a value greater than t_c, while the opposite is true for $b > 1$. The critical case $b = 1$ is simpler, since $t_* = t_c$ and the expression on the RHS of (3.2.10) is linear in t_1.

Strategies on the curve defined through (2.3.94) can resist invasion against mutants that are better/worse defended if they satisfy the convexity condition (2.3.95), which in conjunction with (3.2.3) read

$$\left(\frac{k}{1 + kt_1^*}\right)^2 - \frac{a}{t_1^* - t_c}\left(f - \frac{k}{1 + kt_1^*}\right) > 0. \tag{3.2.11}$$

Under the assumption that $t_1^* > t_c$ this amounts to

$$t_1^{*2} + \left(-\frac{1}{af} - \frac{1}{f} + \frac{2}{k}\right)t_1^* - \frac{1}{fk} + \frac{1}{k^2} + \frac{t_c}{af} < 0. \tag{3.2.12}$$

Since the LHS of (3.2.12) consists of a concave-up, second-order polynomial in t_1 the inequality can only be satisfied if the parabola has two distinct real roots and $t_1^* \geq 0$ is chosen to lie between these. Indeed, we require that the discriminant of the parabola in (3.2.12) is strictly positive, which amounts to

$$a^2 + 2a(1 - 2b) + 1 > 0. \tag{3.2.13}$$

Once more we deduce that stability in the t-direction is conditional on the sign of a certain second-order

polynomial. Indeed, the parabola in a on the LHS of (3.2.13) is concave-up and there are three cases to consider depending on whether its discriminant is negative, zero or positive. The discriminant of the polynomial in (3.2.13) equals $16b^2(1-1/b)$, which suggests that if $b<1$, all choices of $a\in[0,1]$ give rise to a stable region through (2.3.85). If $b=1$ all values of $a\in[0,1)$ will yield a stable region, while for $b>1$ values of $a\in[0,a^-(b))$ work - see Figure 3.1(a). Notice that we have labelled

$$a^{\pm}(b):=2b-1\pm 2b\sqrt{1-1/b}\quad\text{for}\quad b>1, \tag{3.2.14}$$

as the roots of the polynomial in (3.2.13). Values of a that yield a stable region should in principle also be drawn form the interval $(a^+(b),1]$, but it is clear that $a^+(b)>1$ for $b>1$. Furthermore, the smaller root $a^-(b)$ is decreasing over $b>1$ with $a^-(1^+)\approx 1$ and $a^-(b)\approx 0$ for $b\gg 1$. It is also self-evident that no strategy satisfying the equilibrium condition (3.2.2) can be stable in the t-direction whenever $b>1$ and $a\in[a^-(b),1]$. In conclusion, given an appropriate choice of f,k,t_c and α the value of $b=f(1+kt_c)/k$ is such that when it is below unity t-stable strategies can resist invasion against mutant groups of any size, whereas if it is above unity t-stable strategies can withstand invasion against mutant groups of maximum size $a^-(b)$. Arguably, choices of f,k and t_c corresponding to $b=1$ are non-generic.

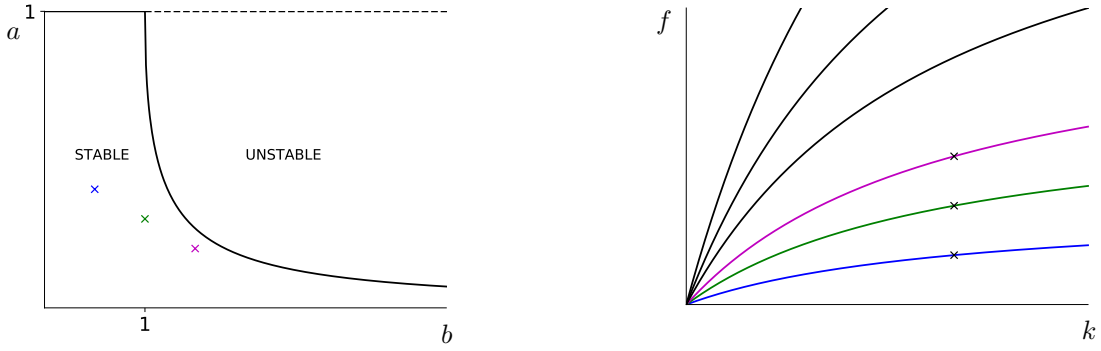

Figure 3.1: On the left figure 3.1(a) are shown the sizes a of mutant groups, against which the residents can withstand invasion (along the t-direction) for given choice of b. A t-stable strategy with f,k and t_c giving $b\geq 1$ have a maximum associated mutant group size of $a^-(b)$ as described in (3.2.14) - which decreases down to zero as b increases - whereas t-stable strategies with $b<1$ can withstand invasion against mutant groups of all sizes. The three x marks with colours blue, green and magenta have coordinates (b,a) given by $(0.5,0.4),(1,0.3)$ and $(1.5,0.2)$ respectively. Figure 3.1(b) on the right shows the b-level curves of the map $(f,k)\mapsto f\big(\frac{1+kt_c}{k}\big)$, with $t_c=0.5$. Blue, green and magenta correspond to $b=0.5$, $b=1$ and $b=1.5$ (as do the x marks in (a)) and the x marks in this figure are positioned at coordinates $(f,k)=(0.5,2)$, $(f,k)=(1,2)$ and $(f,k)=(1.5,2)$.

In particular, we have established that for the appropriate choices of a and b described above - see Figure 3.1 - a strategy (r_1^*,t_1^*) satisfying (3.2.2) satisfies (2.3.92)/(2.3.95) providing that $T^-_{STAB}<t_1<T^+_{STAB}$, where

$$T^{\pm}_{STAB}=\frac{1}{2f}\left(\frac{1}{a}+1\right)-\frac{1}{k}\pm\frac{1}{2af}\sqrt{a^2+2a(1-2b)+1} \tag{3.2.15}$$

are the roots of the polynomial on the LHS of (3.2.12). These roots provide upper and lower bounds to region defined through inequality (2.3.92)/(2.3.95) and are realised as horizontal lines(as they do not depend on r_1) in the strategy space. In particular, points (r_1^*,t_1^*) on the curve (3.2.2) with $T^-_{STAB}<t_1<T^+_{STAB}$ are stable against invasion from mutants (of maximum group size determined by b) that are more or less toxic, while those outside of this region are unstable.

It can be shown that the level of the upper bound T^+_{STAB} shrinks as the proportion of mutants (a) in the local area grows. Indeed the partial derivative of this bound with respect to parameter a gives

$$\partial_a T^+_{STAB} = -\frac{1+\sqrt{p(a)}}{2a^2 f} + \frac{1+a-2b}{2af\sqrt{p(a)}}$$

where we have used the shorthand

$$p(a) := a^2 + 2a(1-2b) + 1.$$

Clearly, $p(a)$ is negative for choices $b > 1$ and $a \in [0, a^-(b))$, while also for the non-generic case $b = 1$ and $a \in [0, 1)$. For choices $b < 1$ and $a \in [0, 1]$ we consider the re-scaled inequality

$$2a^2 f\sqrt{p(a)}\partial_a T^+_{STAB} < 0, \tag{3.2.16}$$

which is equivalent to the trivially-true inequality

$$4a^2b^2\left(1-\frac{1}{b}\right) < 0. \tag{3.2.17}$$

We implement similar reasoning for the lower bound T^-_{STAB}

$$2a^2 f\sqrt{p(a)}\partial_a T^-_{STAB} < 0 \;\Leftrightarrow\; 4a^2b^2\big(1-1/b\big) < 0 \tag{3.2.18}$$

to conclude that for $b < 1$ and $a \in [0, 1]$ the term T^-_{STAB} is shrinking with growing mutant group size a, while it increases for growing a, whenever $b > 1$ and $a \in [0, a^-(b))$. For non-generic choices $b = 1$ and $a \in [0, 1)$ the term T^-_{STAB} is constant over all mutant group sizes. Our observations on the bounding region $T^\pm_{STAB}$ are summarised in Figure 3.2.

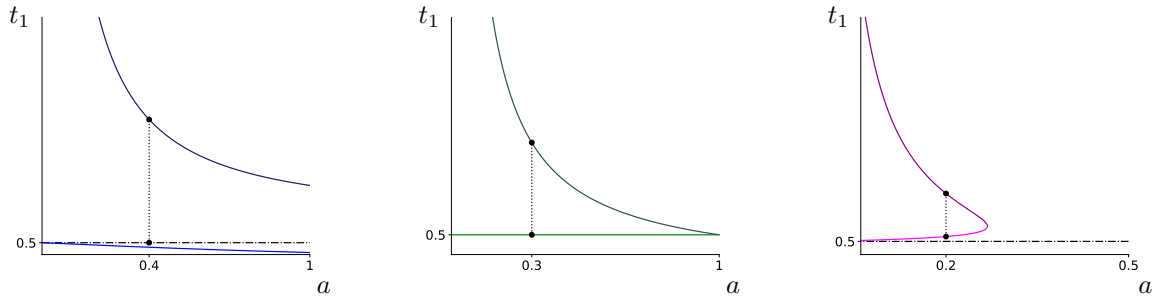

Figure 3.2: Plots showing ${T_{STAB}}^+$ (darker colour shades - top branches) and ${T_{STAB}}^-$ (lighter colour shades - bottom branches) versus a are shown in accordance with (3.2.15), where $t_c = 0.5$ (dashed horizontal lines) has been used. The specific values of b used in Figures 3.2 (a), (b) and (c) have been generated using the values of f and k that are indicated by the x marks in Figure 3.1(b). The x marks in Figure 3.1(a) indicate the values of a at which the dotted vertical line segments are drawn (their lengths indicating the width of the stable bands in Figure 3). Notably the monotonicities of the curves are in agreement with conditions (3.2.16) and (3.2.18). In particular, Figure 3.2(a) contains curves ${T_{STAB}}^\pm(a) = \frac{1}{a} + 0.5 \pm \frac{1}{a}\sqrt{a^2+1}$, line drawn at $a = 0.4$. Figure 3.2(b) shows ${T_{STAB}}^\pm(a) = \frac{1}{2a} \pm \sqrt{a^2-2a+1}/2a$, with ${T_{STAB}}^- = t_c$, dotted line drawn at $a = 0.3$. Figure 3.2(c) shows $\frac{1}{3a} - \frac{1}{6} \pm \frac{1}{3a}\sqrt{a^2-4a+1}$ with the dotted line drawn at $a = 0.2$.

Strategies with non-zero signalling component $r_1 > 0$ are stable in the r-direction providing they can resist invasion against mutants that are either less or more conspicuous. This is guaranteed through conditions

(2.3.92)/(2.3.96) and (2.3.93)/(2.3.97), which in terms of the functional forms in (3.2.1) amounts to

$$-\frac{D'(r_1)}{D(r_1)} + 2\alpha D(r_1)(t_1 - t_c)^2 \left[D'(r_1) + v\left(\frac{1}{a} - 1\right) D(r_1) \right] > 0 \tag{3.2.19}$$

and

$$-\frac{D'(r_1)}{D(r_1)} + 2\alpha D(r_1)(t_1 - t_c)^2 \left[D'(r_1) - v\left(\frac{1}{a} - 1\right) D(r_1) \right] < 0. \tag{3.2.20}$$

We should remark that for cryptic strategies with $r_1 = 0$ only (3.2.20) is required to hold, since these types risk invasion only from the more conspicuous morphs.

3.3 Explicit examples of evolutionarily stable outcomes: A numerical analysis

In this subsection we make use of all results established thus far by assigning specific functional form to the detection rate $D(r)$ and choosing parameters within the partitions introduced in the previous subsection. In particular, we consider

$$D(r) = \frac{d_0}{d_0 + (1 - d_0)\exp(-r)} \tag{3.3.1}$$

and without much loss in generality take $d_0 = 1/2$ - this also suggests that completely cryptic prey that are encountered run a 50% chance of being detected. In addition, we set the critical toxicity level at $t_c = 0.5$. It remains for us to constrain the remaining parameters α, a and v so as to explore a suitable range of admissible solutions. Henceforth, specific plots are produced by making the appropriate choice of parameters for a limited set of equalities and inequalities. From (2.3.94) through to (2.3.97) it is clear that local ESSs consist of subsections of the curve satisfying the equilibrium condition (2.3.94). Before determining explicit outcomes of ESS as subregions of the continuum we utilise the root-finding process presented in (3.2.9) and (3.2.10) to account for the complete breadth of equilibria, all the while maintaining the distinction $b < 1$, $b = 1$ and $b > 1$ - see Figures 3.3 and 3.4). The reader is directed to Figures 3.3 and 3.4 for a demonstration of the root-finding process that is described above. For fixed levels of prey-sensitivity (and associated parameter values) the free parameter is the predation pressure (realised through α) and it is curious to observe its impact on the solutions (r_1^*, t_1^*) of (3.2.2).

Direct substitution of the detection rate (3.3.1) into condition (3.2.2) for (r_1^*, t_1^*)

$$-f + \frac{k}{1 + kt_1^*} + 2\alpha \frac{t_1^* - 0.5}{(1 + \exp(-r_1^*))^2} = 0 \tag{3.3.2}$$

and is generated by interpreting it as an intersection of the LHS with the RHS in the equality below

$$(1 + \exp(-r_1^*))^2 = \frac{2\alpha}{f} \frac{(t_1^* + 1/k)(t_1^* - 0.5)}{t_1^* - t_*}. \tag{3.3.3}$$

The root-finding process of (3.2.10) and (3.3.3) is illustrated directly in the plots of Figure 3.3 and consists of a bifurcation in parameter α. Intersections of the LHS with the RHS in (3.3.3) generate the curve (3.3.2), whose bifurcations in the level of predation pressure α are seen in Figure 3.4 and should be viewed in tandem with Figure 3.3. It should also be noted that although the blue shades have been reserved for $b < 1$, green ones for $b = 1$ and magenta ones for $b > 1$ the parameters of α used in Figure 3.3 and those

used in Figure 3.4 are not the same. In Figure 3.3, solutions to (3.3.3) can be realised as intersections of the coloured curves and the horizontal lines $LHS = c$ with $c \in (1, 4]$. Intersections with $c = 4$ correspond to cryptic solutions with $r_1^* = 0$ while those with $c = 1^-$ correspond to solutions of (3.3.2) that are very bright such that $r_1 \gg 1$. Intersections with ever-decreasing values of c in this interval smoothly generate t_1^* parametrized with increasing r_1 (an explicit formulation in the general setting can be seen in (3.3.2)). The three plots are indicative of the partition-sensitive behaviour discussed thus far and correspond to choices of parameters indicated in Figure 3.1.

As is clear from Figures 3.3 and 3.4 the $b < 1$ regime is especially interesting and there are two critical values for the parameter α, which we define here. The root-finding process is summarised in equalities (3.2.9), (3.2.10), the RHS of which admit a local minimum. The critical values of α are thus defined with respect to their crossing of the limiting values of the LHS of the same equalities (3.2.9) and (3.2.10). That is, the local minimum crosses the bottom dash-dotted line in Figure 3.3(a) for a value of $\alpha = \alpha_1$ defined by

$$-f + \frac{k}{1 + kT_{SIGN}^+} + 2\alpha_1 D^2(\infty)(T_{SIGN}^+ - t_c) = 0. \tag{3.3.4}$$

The same minimum crosses the upper (at value four) dash-dotted line for α given by

$$-f + \frac{k}{1 + kT_{SIGN}^+} + 2\alpha_2 D^2(0)(T_{SIGN}^+ - t_c) = 0. \tag{3.3.5}$$

From these definitions it is clear that there are three different types of solution to (3.3.2) depending on which partition α is chosen from $[0, \alpha_1) \cup [\alpha_1, \alpha_2] \cup (\alpha_2, 1]$.

In Figures 3.3(a) and 3.4(a) the two critical values of α are given by $\alpha_1 = (3 - 2\sqrt{2})/4$, $\alpha_2 = (3 - 2\sqrt{2})$. For $\alpha < (3-2\sqrt{2})/4$ the aversive maximum lies below unity (lower dash-dotted line) and the curve (3.3.2) in Figure 3.4(a) consists of two disjointed roots in the aversive region that come closer with increasing r_1. For $(3 - 2\sqrt{2})/4 < \alpha < 3 - 2\sqrt{2}$ the minimum of Figure 3.3(a) lies between unity and four and so curve (3.3.2) in Figure 3.4(a) exhibits an r_{max}-type vertex in the aversive region. For $3 - 2\sqrt{2} < \alpha \leq 1$ the minimum is above four, which is why in Figure 3.4(a) curve (3.3.2) is not defined.

As explained earlier, the b-partitioning of the parameter space reflects prey sensitivity to investment in defence, such that two distinct regimes are understood: prey with $b < 1$ are t-insensitive, while prey with $b > 1$ are t-sensitive ($b = 1$ is a non-generic description that can be explained using the remaining cases). This means that for given level of investment in toxicity the latter benefit less on account of a comparatively lower rate of reproduction (lower f) and/or a lower level of protection against potentially-lethal attacks (higher k). The differences between the two regimes are manifest in the equilibrium curves (3.3.2), which we explore by varying the level of predation pressure (see Figure 3.4) in these. Comparing Figures 3.4(a) and 3.4(c) one observes that while t-insensitive prey exhibit both increasing and decreasing correlation among aposematic traits (providing predation pressure is moderate-high), t-sensitive prey with $b > 1$ allow for negatively correlated traits only.

A related observation is that when predation pressure is high, $b < 1$ individuals that are moderately aversive tend to advertise increasing levels of defence, while their highly-aversive counterparts tend to advertise reduced levels of toxicity - see disjointed roots in Figure 3.4(a). As predation pressure drops (not below moderate) those moderately-defended prey advertise defence investment up to a certain point (r_{max}-type vertex) beyond which increased toxicity is advertised less. This implies that as predation pressure shifts from high to moderate the range of appearances shrinks so that highly conspicuous morphs become less and less

likely. The effect is compounded as levels of predation pressure drop below moderate such that t-insensitive prey have no conspicuous solutions (recall strategies with $t_1^* < t_c$ fail condition (2.3.93)/(2.3.97)).

This phenomenon makes sense since the only purpose of increased conspicuousness is to reduce the probability of attack (given detection has occurred); as the threat of predation drops this may not be worth increased rates of detection. This should be contrasted with those t-sensitive prey that exhibit aposematic solutions for all levels of predation pressure. The situation is simpler for $b > 1$ individuals for which it holds that as predation pressure drops a certain level of conspicuousness is associated with ever-decreasing (not below the aversive) levels of defence. Once more, this is a sensical effect since for t-sensitive individuals investment in defence is inherently not so beneficial and can therefore only be justified if threat due to predation is considerable. As the latter diminishes, so does the need for investment in defences. A final remark is that when predation pressure is high (regardless of sensitivity to investment in toxicity) evolutionarily stable outcomes are most likely aversive.

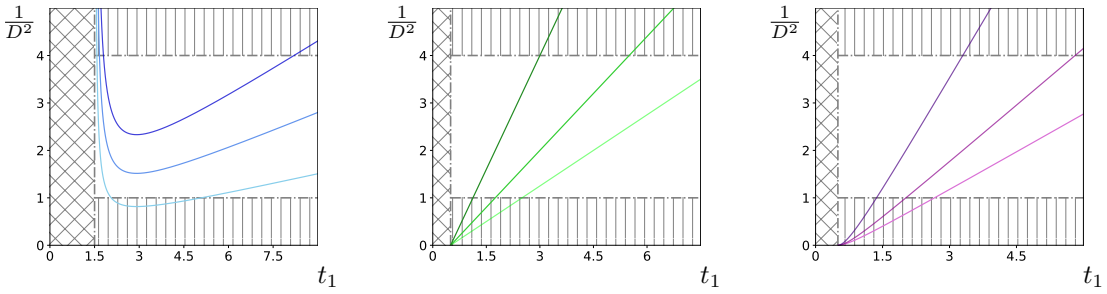

Figure 3.3: Plots show variation in α of the RHS of (3.2.9) for the regimes $b < 1$ (blue), $b = 1$ (green) and $b > 1$ (magenta), such that in each plot the darker shades correspond to higher values of α. Only sections of curves contained inside the unshaded region correspond to real ESSs and there are two grey-shaded regions that are empty of solutions: (i) the diagonally-checkered vertical strips, on which the RHS of (3.2.10) is negative; (ii) the vertically-striped horizontal sections, on which the LHS of (3.2.10), i.e. $1/D^2(r_1)$ assumes values outside of the admissible range $(1, 4]$. (a) Values $f = 0.5$, $k = 2$, $t_c = 0.5$ and $\alpha = 0.035, 0.065, 0.1$ have been used; curves with $\alpha < 0.042$ have minima inside the bottom grey region, curves with $0.042 \leq \alpha \leq 0.171$ have minima inside the unshaded region and curves with $\alpha > 0.171$ are entirely contained in the top grey-shaded region - see also (3.3.4) and (3.3.5). (b) Values $b = 1$, $f = 1$, $k = 2$, $t_c = 0.5$ and $\alpha = 0.25, 0.4, 0.8$ are used and curves with $\alpha > 4$ (not shown) would be entirely outside the unshaded region. (c) Values $b = 1.5$, $f = 1.5$, $k = 2$, $t_c = 0.5$ and $\alpha = 0.4, 0.6, 1.2$ have been used and all values of α yield solutions.

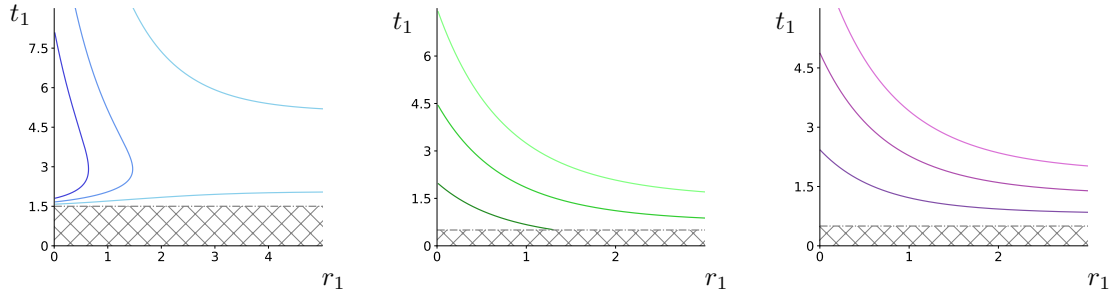

Figure 3.4: Plots show curves of (3.3.2) that result from the root-finding process illustrated in Figure 3.3, such that like-coloured curves correspond to outcomes with equal parameter values. For instance, the bottom curve in Figure 3.3(a), with the lightest shade of blue corresponds to the right-most equilibrium curve of Figure 3.4(a) with the same colour-shading. The diagonally-checkered vertical strips of Figure 3 are seen here as horizontal grey-shaded strips. In Figure 3.4(a) the equilibria are disjointed for small values of α and join together at a local r_{max}, which shifts toward lower conspicuousness until undefined at $r_1 < 0$. In Figures 3.4(b) & (c) The level of toxicity t_1^* satisfying (**??**) shrinks with increasing levels of r_1^* and α.

Having explored the full breadth of equilibria in Figure 3.4, we presently provide examples of these that are locally evolutionarily stable. Sections of the curve (3.3.2) with $t_1^* > 0$ that satisfy (2.3.92)/(2.3.95) if they are contained within $T^-_{STAB} < t_1^* < T^+_{STAB}$, with $T^{\pm}_{STAB}$ given as in (3.2.15). Strategies (r_1^*, t_1^*) with non-zero signalling component satisfy (2.3.96) and (2.3.93)/(2.3.97) if in addition these are contained within the regions bounded by inequalities (3.3.6) and (3.3.7). With parameter values as specified these amount to the pair

$$-\frac{1}{1+\exp(r_1)} + 2\alpha \frac{(t_1-0.5)^2}{(1+\exp(-r_1))^2}\left[\frac{\exp(-r_1)}{(1+\exp(-r_1))} + v\left(\frac{1}{a}-1\right)\right] > 0 \tag{3.3.6}$$

$$-\frac{1}{1+\exp(r_1)} + 2\alpha \frac{(t_1-0.5)^2}{(1+\exp(-r_1))^2}\left[\frac{\exp(-r_1)}{(1+\exp(-r_1))} - v\left(\frac{1}{a}-1\right)\right] < 0. \tag{3.3.7}$$

The discussion thus far has concerned the evolutionary stability of strategies that are aversive (i.e. $I_1 > 0$) and for good reason; these map out the largest region in the strategy space and therefore consist of the overwhelming majority of cases to be considered. Indeed, since the conspicuous signalling of non-aversive strategies is not evolutionarily stable on account of (3.2.19) the only possibility for a non-aversive strategy to be ESS is crypsis. That is, we now briefly account for those strategies with $r_1 = 0$ and $0 \leq t_1 \leq t_c$. The evolutionary stability of aversive strategies has been accounted for by means of the $Q(I_1) = q_0 \exp(-qI_1)$ branch of the attack probability function in (3.2.1), for non-aversive strategies the simpler branch $Q(I_1) = q_0$ is employed for all levels of aversiveness $I_1 \leq 0$.

Alongside cryptic strategies that are aversive, there are now two additional possibilities for a non-aversive cryptic solution: either $(0, t_1^*)$ with $0 \leq t_1^* \leq t_c$ or the origin $(0, 0)$, but not both. On account of (3.2.1) we have

$$Q(I_1) = q_0 \text{ and in particular that } \frac{Q'(I_1)}{Q(I_1)} = 0 \text{ for all } I_1 \leq 0.$$

This assumption suggests that predators do not distinguish between different levels of defence among prey that they find attractive to consume and attack these with some fixed rate. While this assumption is crude it is perhaps not too far from the truth. Furthermore, an immediate consequence is that marginally better (or less)-defended mutant types are attacked by the predators visiting their site at the same rate as residents.

Since Q'/Q features strongly in the conditions for ESS - see (2.3.94) through to (2.3.97) these are notably simpler here. To that, we observe immediately that stability in the r-direction for non-aversive cryptic strategies is guaranteed automatically since (2.3.87) amounts to the trivially-true inequality

$$-\frac{D'(0)}{D(0)} < 0. \tag{3.3.8}$$

The latter is straightforward; since detectability is the only factor influencing marginal differences in fitness the more conspicuous mutant admits a clear disadvantage compared with the resident. It also follows that the evolutionary stability of non-aversive strategies is governed by their stability along the t-direction.

The origin is a local ESS if (2.3.86) holds. Setting $\lambda = 0$ and $Q'/Q = 0$ into (3.1.6) suggests that this is true if $f > k$. Strategy $(0, t_1^*)$ with $t_1^* < t_c$ must satisfies (2.3.91), which is true if

$$-f + \frac{k}{1 + kt_1^*} = 0 \;\; \Leftrightarrow \;\; t_1^* = \frac{1}{f} - \frac{1}{k}. \tag{3.3.9}$$

As it happens, any solution $(0, t_1^*)$ of (3.3.9) satisfies inequality (2.3.92) since setting $Q'/Q = 0$ into this amounts to the trivially true statement

$$-f^2 + \frac{2k^2}{(1 + kt_1^*)^2} > 0 \;\; \Leftrightarrow \;\; t_1^* < \frac{\sqrt{2}}{f} - \frac{1}{k}. \tag{3.3.10}$$

In summary, we deduce that if $f > k$ then the origin is a non-aversive ESS, while if $f < k$ and $1/f - 1/k < t_c$ then the LHS of the latter is an ESS.

We have mentioned that the surface provided in the LHS of (3.2.3) consists of a family of concave-up parabolas parametrized by r_1 such that the equilibrium curve consists of the points in the strategy space where it intersects the (r_1, t_1)-plane. With regards to this picture, the partitioning of the parameter space is as follows. If $b < 1$ the family of polynomials crosses the (r_1, t_1)-plane twice (in approaching and receding the minimum surfaces local t_1-minimum). As r_1^* increases the associated level of defence for the top root sinks while the defence for the bottom root crossing increase until for some critical level of the signalling the two roots meet - see Figure 3.5(a). The case $b = 1$ describes the borderline case (non-generic) in which the surface of polynomials in (3.2.3) doesn't cross but "touches" the (r_1, t_1)-plane along a family of local t_1-minima shown in Figure 3.5(b); if $b > 1$ the surface crosses the (r_1, t_1) - plane at most once for given r_1 at a level that corresponds to the top root - the other root is non-sensical - which naturally decreases with increasing r_1 as shown in Figure 3.5(c).

Figure 3.5 contains important examples of ESSs, which constitute distinct realisations emerging from the choices for parameters a, b, f, k and t_c that are indicated by the x marks in the plots of Figure 3.1; additional choices for a and v are specified in the caption. In Figure 3.5(a), strategies in the region below the bottom brown curve fail condition (3.3.6), while strategies above the top brown curve fail condition (3.3.7), which implies that the section of (3.3.2) that lies between these two curves satisfies both (3.3.6) and (3.3.7). The subsection that is also contained within the blue solid lines (as indicated by the solid markers) satisfies $T_{STAB}^- < t_1^* < T_{STAB}^+$ with $T_{STAB}^{\pm}$ given as in (3.2.15) and therefore contains local ESSs, whereby traits can either be positively or negatively correlated. There are two intersections of (3.3.2) with the vertical axis $r_1 = 0$. The first is not shown and is not a cryptic ESS as it has $t_1^* < T_{STAB}^-$, while the other intersection (shown) is a true cryptic ESS $(0, t_1^*)$, since it satisfies (3.3.7) and $T_{STAB}^- < t_1^* < T_{STAB}^+$, i.e. it sits below the top brown curve and between the two solid blue lines. It follows directly that the origin is not a cryptic

ESS.

Figure 3.5(b) is more straightforward to analyse because (3.3.7) is everywhere satisfied. The pair of brown dash-dotted lines are the zero level sets of the LHS of (3.3.6), so that only the region inbetween these fails the condition in (3.3.6). Since the curve in (3.3.2) is entirely above the top brown curve, the section of this that is below the top solid green line (as indicated by the marker) consists of local ESSs, all of which suggest a negative correlation between conspicuousness and defence. The intersection of (3.3.2) with $r_1 = 0$ is given as $(0, t_1^*)$ with $t_1^* > T_{STAB}^+$ and is therefore not an ESS as it fails (2.3.92); the origin $(0, 0)$ is a true cryptic ESS.

In Figure 3.5(c) the top brown dash-dotted curve is the zero level set of the LHS of (3.3.7), so that the region above it fails the condition in (3.3.7), while the pair of brown curves below this constitute the zero level set of (3.3.6) so that the region inbetween them fails inequality (3.3.6). This means that the section of the curve (3.3.2) in between the top and middle brown satisfies both (3.3.6) and (3.3.7) and therefore contains local ESSs; the next section (as indicated by the solid markers) below the middle brown curve fails (3.3.6) and the last section of the curve above the brown curve again satisfies both (3.3.6) and (3.3.7) and therefore contains local ESSs. The intersection of the curve (3.3.2) with $r_1 = 0$ is a cryptic ESS $(0, t_1^*)$ since it satisfies $T_{STAB}^- < t_1^* < T_{STAB}^+$ and (3.3.7). The origin is not a cryptic ESS.

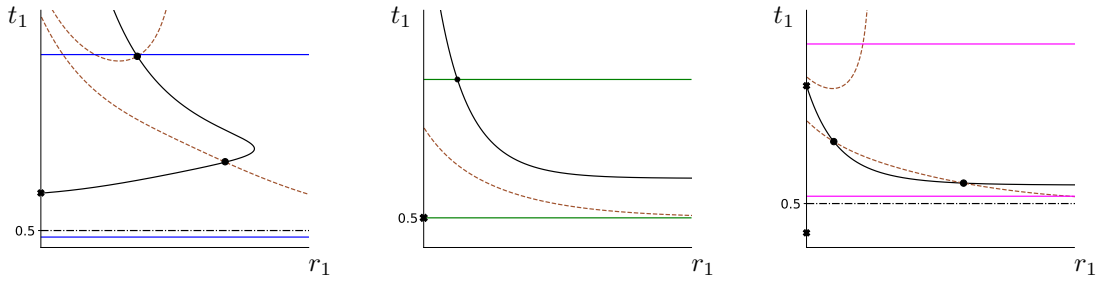

Figure 3.5: All three figures contain plots of local ESSs in strategy space as subsections of the t-equilibrium curve (3.3.2) shown as black solid lines and partitioned by black markers (filled X marks indicate cryptic solutions). The dashed-dotted lines mark the natural cut-off at $t_c = 0.5$, while the brown dashed lines represent the zero level sets of the LHS of (3.3.6) and (3.3.7). In (a) we have set $b = 0.5, a = 0.4, f = 0.5, k = 2, \alpha = 0.0445$ and $v = 0.04$. The solid curve has an r_{max}-type vertex at $t_1 = T_{SIGN}^+ = 1.5 + \sqrt{2}$ and unique cryptic ESS at the aversive level $t_1 \approx 1.601$. In (b) we have set $b = 1, a = 0.3, f = 1, k = 2, \alpha = 0.3$ and $v = 0.4$. The section of the t-equilibrium curve that lies under the upper green solid line and above the brown dashed line consists of a continuum on which the ESS level of defence decreases with increasing level of the conspicuousness; $t_1 = 1/2$ is the unique cryptic ESS, although this may be non-generic. In (c) we have set $b = 1.5, a = 0.2, f = 1.5, k = 2, \alpha = 1.6$ and $v = 0.05$. The section of the solid black curve in between the lower and upper brown dotted lines consists of a decreasing continuum of local ESSs, the section under the lower brown curve does not contain ESSs, while the last section that is once more in between the middle and top brown lines consists of a second decreasing continuum of ESSs. The intersection of the equilibrium curve with $r_1 = 0$ is an aversive cryptic ESS, while $t_1 = 1/6$ admits a non-aversive cryptic solution.

3.4 Remarks on the convergence stability of local ESSs

A trivial re-arrangement of (3.3.2) reads

$$\frac{k}{1 + kt_1^*} + 2\alpha D^2(r_1^*)(t_1^* - t_c) = f. \tag{3.4.1}$$

The two terms on the LHS account for the marginal advantage of the more conspicuous mutant associated with (i) a reduced probability of being captured and (ii) a reduced rate of attack (perceived as more aversive).

The term on the RHS represents the marginal disadvantage of the more toxic mutant associated with reduced fecundity. A resident strategy (r_1^*, t_1^*) is an element of (3.4.1) if the LHS equals the RHS, or rather, to first order, the marginal advantage of the better-defended mutant exactly matches it's fecundity disadvantage. The mutant fitness is almost everywhere $\mathcal{C}^l$ for $l \geq 2$ and therefore the argument can be reversed when analysing the trade-off to a less-defended mutant. For $(r_1^*, t_1^*) \in \mathcal{D}_2 \sqcup \mathcal{D}_3$ to be locally evolutionarily stable we also require that it satisfies (2.3.92)/(2.3.95). Physically, this condition guarantees that more/less toxic mutants ultimately do experience a fitness disadvantage compared with their resident counterparts but these enter in as second order terms.

Strategies failing the equilibrium condition (3.4.1) in such a way that the LHS is less than the RHS are situated on a *down-slope* of the mutant payoff along t_1 in the sense that the marginal fitness advantage of the better-defended mutant is less than its fecundity disadvantage. Increasing the mutant defence in the vicinity of such a resident strategy leads to a drop in the mutant fitness. Conversely, resident strategies for which the LHS is greater than the RHS in (3.4.1) are on an *up-slope* of the mutant payoff along t_1 in the sense that increasing the mutant toxicity in the vicinity of the resident strategy will increase its marginal advantage more than fecundity disadvantages will decrease it.

It is clear that the LHS of (3.4.1) is increasing with respect to r_1^* on account of the term proportional to D^2; its monotonicity with respect to t_1^* however is less straightforward. The LHS is t_1^*-increasing if and only if

$$\partial_{t_1^*} LHS > 0 \quad \Leftrightarrow \quad 2\alpha D^2(r_1^*) - \frac{k^2}{(1+kt_1^*)^2} > 0. \tag{3.4.2}$$

As we established in our discussion of 3.1.1, the above inequality is satisfied whenever $t_1^* > T_{SIGN}^+$, whenever $b < 1$ and is trivially true over the aversive and conspicuous region whenever $b \geq 1$. Indeed, this provides a more tangible explanation of why the curve in (3.2.4) in these instances is understood as the graph of a function $t_1^*(r_1^*)$ whose value decreases with increasing levels of the conspicuousness r_1^*. Since the LHS of (3.4.1) is increasing with respect to r_1^* this means that for a marginal increase in r_1^* the associated quantity t_1^* decreases (by an amount that depends on the slope of the curve) so as to compensate and keep the LHS constant and equal to the RHS, i.e. f. The only instance in which the curve defined through (3.4.1) is increasing is for $t_1^* < T_{SIGN}^+$, whenever $b < 1$. In this case, the reverse of inequality (3.4.2) holds, which means that the LHS of (3.4.1) is decreasing with respect to t_1^*. A direct implication of this is that an increase in r_1^* must be coupled by an increase in t_1^* to maintain equilibrium.

We now restrict our attention to regions of the strategy space in which both (3.3.6) and (3.3.7) hold, which can be achieved by picking large enough v. We can use the above reasoning to determine which subsets of the equilibrium continuum are convergence stable and which are not. The general result is that sections of the continuum on which conspicuousness and defence exhibit a decreasing relationship are unstable, while those on which the relationship is increasing are stable. Stability in this instance is understood as follows. We begin by considering the instance $b < 1$ and the branches of the curve (3.4.1) separately. Assume that (r_1^*, t_1^*) is contained in the bottom branch of (3.2.4) the resident population is playing some strategy $(\hat{r}_1, \hat{t}_1)$ with $\hat{t}_1 < t_1^*$. Since $\hat{t}_1 < t_1^* < T_{SIGN}^+$ it follows that the LHS of (3.4.1) is decreasing with respect to the resident level of defence, implying in turn that the mutant payoff in the vicinity of $(r_1^*, \hat{t}_1)$ is situated on a down-slope of the resident defence. If we hypothesise a process whereby the resident strategy shifts toward increasing fitness through a sequence of selectively advantageous mutations then it is clear that over time resident strategy will tend to the equilibrium strategy (r_1^*, t_1^*).

Making use of the property that the mutant fitness is almost everywhere $\mathcal{C}^l$ with $l \geq 2$ in the vicinity of

the resident value it also holds that if the residents were playing a strategy $(r_1^*, \hat{t}_1)$ with $t_1^* < \hat{t}_1 < T_{SIGN}^+$ then such a strategy would be situated on a down-slope of the mutant payoff along t. By the same token, a sequence of beneficial mutations would send the resident strategy to the equilibrium value vertically and along r_1. In this conception it is also notable that mutations in r cannot invade as we can freely pick sections of the curve that are contained in the region specified by inequalities (3.2.19) and (3.2.20). Convergence stability of the bottom ESS branch in the $b < 1$ description is therefore summarised by a sequence of vertical arrows pointing toward the branch.

Continuing with the $b < 1$ example, we consider the top branch of (3.2.4) and in particular, the convergence stability of a point (r_1^*, t_1^*) on this. Assume that the prey population is playing some strategy $(r_1^*, \hat{t}_1)$, where $T_{SIGN}^+ < \hat{t}_1 < t_1^*$. Since for $\hat{t}_1 > T_{SIGN}^+$ the LHS of (3.4.1) is increasing with respect to t_1^* it is clear that such a resident strategy is on a t down-slope of the mutant payoff. This means that the LHS of (3.4.1) is less than the RHS and that a less toxic mutant has a fitness advantage. This also means that selectively advantageous mutation will send the prey strategy toward the lower branch of the equilibrium curve along the $r_1 = const.$ vertical line. Strategies between the two branches are shifted toward the bottom branch under this process. It is also true that if the prey were to play some strategy $(r_1^*, \hat{t}_1)$ with $\hat{t}_1 > t_1^*$ that the LHS of (3.4.1) would exceed the RHS, implying that the strategy is situated on a t up-slope of the mutant payoff. As is expected from the differentiability of the mutant payoff along t it is the case that successive mutation along the direction of increased fitness will eventually send the population toward ever-increasing values of toxicity along the vertical $r_1 = const.$ line.

In particular, we have established that for $b < 1$ the top branch of the equilibrium curve is not convergence stable and indeed, in an r- stable vicinity, neighbouring strategies with higher toxicity are "repelled" along a vertical line, while strategies with lower toxicity are "repelled" along vertical lines that point toward the bottom branch. The bottom branch is convergence stable in the sense that neighbouring strategies in an r-stable vicinity with toxicity greater or lower than the equilibrium value are "attracted" toward the branch along vertical lines. The convergence stability analysis described above can be extended easily to the equilibrium curves with $b \geq 1$. Indeed, in such instances the curve of (3.4.1) is manifest as a decreasing toxicity-conspicuousness continuum that is convergence unstable in the same sense that the top branch of the $b < 1$ equilibrium curve is convergence unstable. Furthermore, this also means that the continuum (on which conspicuousness and defence are increasing) described by the example functions of Broom et al. (2008) is also convergence stable in the same sense. That is, in the vicinity of a strategy (r_1^*, t_1^*) that satisfies (3.3.6) and (3.3.7) resident strategies are attracted toward equilibrium along vertical lines.

3.5 Some discussion on the honest signalling of aposematism

In this chapter we have built on the game-theoretical model of Broom et al. (2006) in accordance with Scaramangas and Broom (2022), whose findings constitute an important exploration into the far-reaching implications of the former. We have established that Broom et al. (2006) can be released from some restrictive conjectures that were made upon its publication and hence argue that it is the only model-based study to provide sound justification for some more curious instances of aposematic behaviour observed among species of the *Dendrobatidae* family. In particular, we have shown that evolutionarily stable combinations of aposematic traits can be either positively or negatively correlated (on an across-populations basis) and that for given level of conspicuousness there can exist more than one optimal levels of defence. Our results were demonstrated in section 3.3, in which we consider a set of functional forms that lead us to explore

a larger plethora of evolutionarily stable outcomes, in which the relationship between aposematic traits is naturally more complex. The example of (3.2.1) is biologically plausible and the procedure for studying its properties relating to evolutionary stability are indicative of the complexities associated with the stability of natural systems. Indeed, we have discovered that the sensitivity of prey to investment in defence, which is an inherent property and perhaps indicative of the stage of evolution of the defence in question (early-stage defences tend to incur larger costs as suggested in Tarvin et al., 2017) plays an important role in the outcome that we expect to see.

In addition to the game-theoretical treatments of Leimar et al. (1986) and Broom et al. (2006) that are discussed in the introduction, there exists a range of publications that examine the co-evolution of aposematic traits from the mathematical modelling perspective and within these a variety of driving mechanisms and notions for optimality are considered (see Summers et al., 2015 for a systematic review). Among these, the majority (see for example M. P. Speed and Franks, 2014, Holen and Svennungsen, 2012, Lee et al., 2011, Blount et al., 2009, and Franks et al., 2009) alludes to the idea that conspicuousness and defence should be positively correlated, with M. P. Speed and Ruxton (2007) admitting the only exception to this. The latter suggests that this correlation can be either positive or negative depending on the variation in marginal costs of aposematic display across populations. Although sound and intuitive, the mechanism presented therein is not expounded in the level of detail that is used here, nor are strategies 'assessed' in terms of their evolutionary stability. Another contribution is that of Svennungsen and Holen (2007), who investigate the possibility of a stable dimorphism using a game-theoretical description that is similar to our own. However, their focus is automimicry and deal with a single defensive trait as opposed to considering the joint co-evolution of a defensive and signalling trait.

Scaramangas and Broom (2022) is the only theoretical study suggesting that negatively correlated aposematic traits (across different populations) could be evolutionarily stable (as mentioned in the introduction claims that Leimar et al., 1986 identified negatively correlated solutions are not correct). It is important to stress that in our study correlations between aposematic traits are made on an across-populations basis. This is neither true among all model-based studies, nor among empirical studies, some of which additionally consider within population variations while others consider variations across species.

An important assumption of Broom et al. (2006) that is perhaps not emphasised in that publication is that predators learn quickly to avoid prey that are unpalatable and that aversive learning occupies a short investigative period that takes place early on in their life. Insectivorous birds residing in tropical regions may consume hundreds of insects a day and live several years; for the duration of their life they impose mortality based on these early experiences and could closely fit the predator assumptions of Broom et al. (2006) detailed above. There is a considerable volume of evidence in support of this "fast learner" assumption, including the field observations of L. P. Brower (1969) on blue jay - monarch butterfly systems, the commentary of Mallet (2001) and the observations of Darst et al. (2006) among others, on chickens feeding on poisonous *Dendrobatidae* frogs.

Among empirical studies, the majority observe that aposematic traits are positively correlated. In particular, Arenas et al. (2015), Blount et al. (2012), Maan and Cummings (2012a), Vidal-Cordero et al. (2012), Santos and Cannatella (2011), Cortesi and Cheney (2010), Bezzerides et al. (2007) and Summers and Clough (2001b) have all observed positive correlations, while I. J. Wang (2011) and Darst et al. (2006) observed negative correlations and lastly Daly and Myers (1967) observed no correlation whatsoever. These studies have considered samples of taxa ranging from marine opisthobranchs, wasps, beetles and frogs and while some consider within-population variations, others look at variations across populations and others yet consider

different species of a given genus. Prior to Scaramangas and Broom (2022), no model-based approach could provide sound reasoning to support the possibility of negative correlations and indeed we argue that the latter is the only study containing causal confirmation that evolutionarily stable combinations of aposematic traits need not be positively correlated (but instead follow a more complex relationship).

The authors of Broom et al. (2006) had originally anticipated that conspicuousness and defence should be positively correlated. Indeed, this result was confirmed by the functional forms considered therein and subsequently in Broom et al. (2008) and is in fact a sensible assertion to make: not only does this appear to be the prevailing pattern in nature but one would expect that more conspicuous prey should be better defended as they are more likely to be attacked. As we have demonstrated, this reasoning is not entirely accurate, especially once aposematism has become established within a population of prey (although it may better apply during the evolutionary transition from crypsis to aposematic signalling) and instances in nature that appear to negate this cannot be ignored. Indeed, using reflectance spectra and toxicity assays on different populations of *Oophaga granulifera* (a species of poison frog) I. J. Wang (2011) observed that the less conspicuous morphs were the most toxic, while the most conspicuous ones were the least toxic. Darst et al. (2006) observed a similar effect among different genera of the *Dendrobatidae* family and hence hypothesised that aposematic traits become decoupled so that arbitrary combinations can reduce anti-predator attack rates. This suggestion supports our own conclusion that for given level of conspicuousness the optimal level of toxicity (providing this exists) need not be unique. Furthermore, Darst et al. (2006) propose that potentially costly unprofitability may be traded off in favour of bright colouration so that optimal investment in secondary defence will diminish when more cost-effective conspicuousness evolves as a primary defence.

It should be noted that this differential costs analysis of Darst et al. (2006) is based on the following assumptions: both the production of the signal and production of the defence are physiologically costly, even if no attacks occur on a particular prey individual (by contrast, we assume that only investment in toxicity is physiologically costly); increased conspicuousness is both increasingly physiologically costly and increasingly effective in reducing the likelihood that a predator successfully kills discovered prey; the same assumptions are true of increased investment in the defence. Thus there is a natural trade-off between investment in either signal or defence that leads to the reported negative correlation; although this may provide a sufficient explanation, it may in fact not be necessary, since the same effect can be explained by evolutionary stability considerations alone. A concise explanation of how we retrieved negatively correlated solutions in this chapter cannot be given in an equally similar manner; the model at hand is far more elaborate and as has become clear from comparison with Broom et al. (2008) that the particular results retrieved depend strongly on the functional forms chosen. While in Broom et al. (2008) only increasing solutions are possible, the slight modification in Q has opened up the possibility for both positively and negatively correlated traits. Indeed, it is notable herein that individuals that are more sensitive to investment in toxicity ($b \geq 1$) exhibit a negative correlation, while less sensitive individuals ($b < 1$) can have both positive and negative correlations; such an effect is unlikely to hold generally.

As far as empirical testing of the model goes, this is a difficult process; the model is elaborate and controlling all the parameters involved would admit an unrealistic task for the biologist. Nonetheless, the connection between Broom et al. (2006) and Darst et al. (2006) can be clarified to some extent. In particular, although the latter is most concerned with the effectiveness of choice of strategy on reducing attack by predators, we are more concerned with its contribution to overall prey fitness. Fitness is understood as the number of offspring produced per life cycle and naturally depends on predator-induced mortality, suggesting that the work of Darst et al. (2006) is important, but which should be supported by additional demographic

measurements. For example, a population consisting of mainly young prey would indicate high fecundity and predator-induced mortality, while a population consisting mainly of old prey suggests low fecundity and low predation pressure; these data can be cross-examined with strategy considerations by means of toxicity assays and reflectance measurements as in Darst et al. (2006) to establish a connection between strategy and fitness. Such measurements rely on knowledge of average lifespan among camouflaged prey and should be carried out on an across-populations basis; in practice it may be difficult to identify populations of a given species playing a range of different strategies. Finally, it would be interesting to establish whether investment in conspicuousness is physiologically costly; this would require that we find two populations that are equally toxic and differ only in conspicuousness and even so, demographic differences may not be directly related to differences in the fecundity but to the overall fitness instead.

This brief accounting of empirical and model-based studies (and indeed of this present chapter's relevance to these) would be incomplete if some inherent difficulties in interpreting empirical findings were not pointed out. For example (i) studies on certain animals may be naturally more/less relevant to a given model and further, (ii) some are carried out on an intrapopulation basis, others on an across-populations and others yet on an across-species basis (Summers et al., 2015). Notably (i) and (ii) limit the number of potentially relevant studies (there are already not that many) to any given model. (iii) It is believed that there are several factors driving aposematism in real systems (such as variation in life history or community structure for example) and the extent of their individual contributions in any given system may not be generally known (see M. P. Speed and Ruxton, 2007). (iv) Experimentalists use different operational definitions of conspicuousness and toxicity and the specific techniques used for determining these may make it difficult to compare empirical studies to each other, even if the variables in (i) and (ii) are fixed (see chapter 6 of Ruxton et al., 2019).

This chapter and Scaramangas and Broom (2022) constitute a significant advancement to the model of Broom et al. (2006), which to this day is among the leading model-based treatments of aposematism. Nonetheless, there are areas that remain to be explored within this and which we invite the reader to consider. For instance, while the original model does account for the impact of prey appearance and the degree of kin grouping on the evolutionarily stable levels of defence within the population, this is ignored in Scaramangas and Broom (2022) and indeed throughout this manuscript. Indeed, while our omission of the contribution of colouration to prey appearance (and the existence of non-point solutions) was to avoid unnecessary complexity with regard to studying the co-evolution of conspicuousness and defence, it remains an omission, which we encourage the reader to consider together with the new ideas in this chapter.

For example, although we provide explicit examples of local ESSs and discuss in section 3.4 whether these can be attained through small, selectively advantageous mutations (see *convergence stability*) we do not provide explicit calculations on this but rely on verbal explanations. Indeed, it would be of interest to determine whether a population playing an unstable strategy in the locality of the an ESS could eventually converge to the latter. This is precisely the theme of the following chapter.

The model of Broom et al. (2006) can be extended to consider mimicry systems as well as more general co-existence regimes. For example, we may conceive of a scenario wherein the prey population is made up two types (belonging or not to the same species) each playing a different strategy such that the intention is to determine whether these can co-exist in a certain proportion over the long term. This is a possibility if the prey population is in a state of *stable equilibrium*, whereby the more fast-paced population dynamics are stable and in addition, each type is (locally) evolutionarily stable in the sense of the definition provided in section 2. Stable co-existence is an interesting problem to consider in generality and particularly so because

Batesian and automimicry systems are specific examples of this. Indeed, imposing that both types of prey are equally conspicuous and such that one is positively aversive, while the other is attractive describes a mimicry situation. Although the most immediate extension is automimicry, in which both types belong to the same species, our own interest is in Batesian mimicry - this is also more challenging because to each species pertains a different set of functional forms (such as those considered in section 3).The work of Svennungsen and Holen (2007) is particularly relevant to the extension of our own model to automimicry systems; they investigate the possibility of an evolutionarily stable dimorphism within a game-theoretical framework that resembles our own. As mentioned previously, however, our model studies the joint co-evolution of aposematic traits as opposed to aposematic defences in isolation. Finally, whether or not conspicuous signals can be better recollected by predators in not fully understood and so it may be interesting to consider instances in which the rate of recollection is not simply a scalar of the rate of detection, but is instead scaled by a function that increases with prey conspicuousness.

Chapter 4

Theoretical and genetic algorithm predictions on ESSs with non-zero rates of background mortality

Our understanding of aposematism (the conspicuous signalling of a defence for the deterrence of predators) has advanced notably since its first observation in the late nineteenth century. The purpose of this chapter is two-fold: first, to determine the relationship between evolutionarily stable levels of defence and signal strength under various regimes of background mortality and colony size (previous attempts have assumed predation to be the only source of death). Second, to compare these predictions with simulations of finite prey populations that are subject to random local mutation. We consider jointly the roles of absolute resident fitness, marginal mutant fitness and stochasticity in the evolution of prey traits and discuss the importance of population size in the above. The work presented here provides new insight into Wallace's first recorded *"warning colouration"* in animals. Indeed, it extends the scope of the celebrated model by Broom et al. (2006) both from the analytical standpoint (by accounting for regimes of varying background mortality and colony size) and from the practical standpoint (by assessing its efficacy and limitations in predicting the evolution of prey traits in finite simulated populations). Both developments constitute new contributions to the theory of aposematic signalling. The presentation in this chapter follows closely the (submitted) publication by Scaramangas et al. (2022), which the reader is encouraged to consult alongside this.

The observation of aposematism in the natural world would seem troubling from the evolutionary standpoint as it is sensible to surmise that conspicuous individuals run a clear disadvantage compared with their non-signalling counterparts. It would be unfeasible to control all the parameters of the model in a free experimental setup and at the very least, challenging in a laboratory setup; we therefore seek to compare our predictions for a finite population of prey within the context of a genetic algorithm model. We seek to address three areas that appear less acknowledged in the literature: (a) In a large enough (effectively infinite) population of prey is there a certain manner in which defence should be advertised to make the population more likely to retain its composition over successive generations and under the presence of mutation? (b) how might our answer in (a) change under different regimes of background levels of mortality? (c) how might our answers in (a) and (b) be adapted to a population of prey that is finite but large enough that its traits are not fully driven by stochasticity?

4.1 ESS analysis for the genetic algorithm model

In the previous chapter we considered the functional forms used in Broom et al. (2008) with modification in the dependence of Q on the perceived aversiveness - see (3.2.1). While plausible, this modification in Q allowed us to demonstrate that it is not generally possible to arrive at an explicit expression involving the equilibrium level(s) of defence and the conspicuousness. In this section we re-visit the forms of Broom et al. (2008) and introduce non-zero rates of background mortality, which we consider alongside non-zero levels of the local relatedness (a circumstance not previously explored). Presently we predict that (for fixed signalling component) the equilibrium level of defence decreases as the background mortality rate increases. In addition, we predict that an increase in the same parameter may cause a shift from a decreasing relationship to an increasing one. The above results are confirmed numerically in sections 4.3 and 4.4 and are interesting not only from the mathematical modelling perspective but are perhaps of innate interest to the biologist, particularly since there is no clear consensus regarding how different components of aposematic traits are and should be related in the natural world - see Summers et al. (2015).

In this section we outline a general procedure for interpreting the evolution of prey traits in the simulated populations by considering (i) the marginal mutant fitness along the t-direction; the marginal mutant fitness along the r-direction and (iii) the dependence of the absolute resident fitness on the equilibrium conspicuousness. The developments of this section form the building block for the analysis of the simulations in sections 4.3 and 4.4, in which we discuss explicit examples with $\lambda = 0$ and $\lambda > 0$, treating occasions with zero and non-zero colony size separately therein. The functional forms are given as

$$F(t) := f_0 \exp(-ft); \quad H(t) := t - t_c; \quad K(t) := \frac{k_0}{1+kt}$$

$$L(r) = D(r) = \frac{1}{1+\exp(-r_1)}; \quad Q(\mathcal{I}) := min\left(1, q_0 \exp(-q\mathcal{I})\right); \quad S(x) = max(1-vx, 0) \tag{4.1.1}$$

and are shown in Figure 4.1 below. We emphasise that while previous works including Broom et al. (2008) have considered (4.1.1) only alongside $\lambda = 0$ we presently account for scenarios with $\lambda = 0$ and $\lambda \neq 0$ alongside zero and non-zero levels of local clustering.

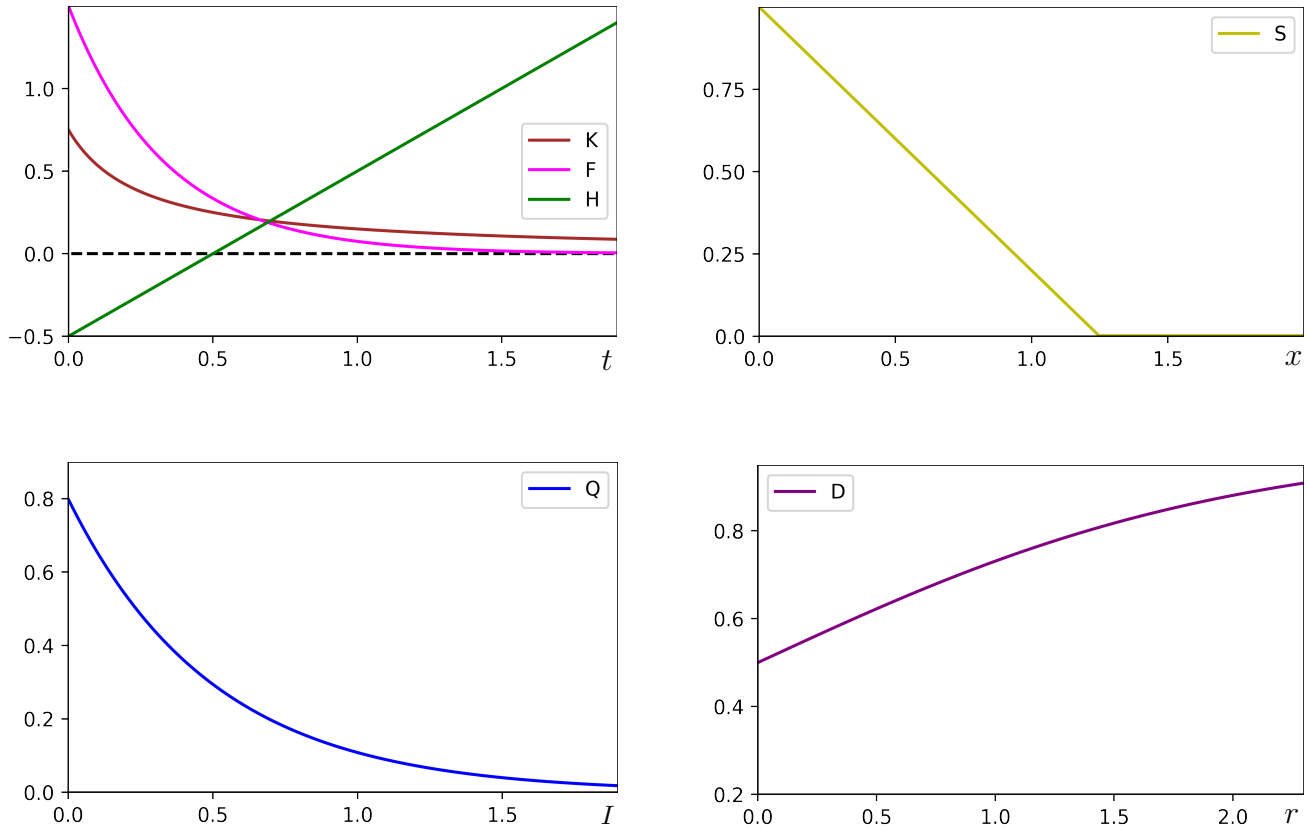

Figure 4.1: The example functions of (4.1.1) as used in the genetic algorithm model with specific parameter values chosen for purposes of demonstration only. **(a)** [Top Left] The functional forms for the probability of escaping a mounted attack (K), the fecundity (F) and the aversiveness of an experience (H) plotted as functions of prey toxicity. Parameter values given as $k_0 = 0.75$ and $k = 4$; $f_0 = 1.5$ and $f = 3$ and $t_c = 0.5$ respectively. **(b)** [Top Right] The functional form for the (uni-variate) similarity function S plotted with respect to the generic variable x and parameter $v = 0.8$. **(c)** [Bottom Left] The form for the probability of attack Q with $q_0 = 0.8$ and $q = 2$ plotted as a function of the perceived aversiveness; **(d)** [Bottom Right] The form for the rate of detection plotted as a function of prey conspicuousness D with $d_0 = 0.5$.

Mutant fitness

We now narrow our attention to the functional forms of (4.1.1) and draw analytical conclusions about the behaviour of the system at ESS. For a strategy $(r_1^*, t_1^*) \in \mathcal{D}_2 \sqcup \mathcal{D}_3$ to be locally evolutionarily stable it must satisfy (2.3.91)/(2.3.94). Through the forms in (4.1.1) equality $g_1(r_1^*, t_1^*) = 0$ reads

$$-\frac{\lambda f}{q_0 k_0}(1 + kt_1^*)(1 + \exp(-r_1^*)) \exp\left(q\frac{N}{n}\frac{t_1^* - t_c}{1 + \exp(-r_1^*)}\right) - f + \frac{k}{1 + kt_1^*} + \frac{aq\frac{N}{n}}{1 + \exp(-r_1^*)} = 0. \tag{4.1.2}$$

There is an immediate conclusion to be drawn from the above, which confirms our intuition that anti-predatory defences are of diminishing value in regimes of increasing non-predatory threat. Indeed, suppose that for some level of background mortality $\lambda = \lambda^*$, the strategy (r_1^*, t_1^*) is a solution to (4.1.2). Since the LHS of that equality decreases with increasing values of either t_1^* and/or λ^* it follows that an increase in λ would lead to a decrease in the equilibrium toxicity t_1^* associated with conspicuousness r_1^*. That is, for fixed conspicuousness, the equilibrium level of defence decreases with increasing levels of the background mortality. We would expect there to be little value in investing in defences that are costly to the fecundity

in regimes where these have limited capacity to increase prey life-span and it is worth mentioning that this result is confirmed in the simulated plots of Figure 4.7 (see end of section 4.4).

It is immediately clear from (4.1.2) that if $\lambda > 0$ and $a > 0$ it is not possible to obtain the ESS level of toxicity explicitly in terms of the conspicuousness. This is unlike the situations encountered previously in Broom et al. (2006) and Broom et al. (2008) and is indicative of a broader class of examples in which one trait can only be determined in terms of the other at ESS through a rule that is implicit. In chapter 3 as well as in Scaramangas and Broom (2022) it is demonstrated that in such cases the relationship between conspicuousness and defence along the curve given through $g_1(r_1^*, t_1^*) = 0$ in (3.1.6) can be better understood through the *Implicit Function Theorem* in $\mathbb{R}^2$. Presently we provide a derivation for the slope of the line tangent to the (implicitly defined) curve by utilising (3.1.11) and identifying the LHS of (4.1.2) with $g_1(r_1, t_1)$. The r_1-derivative of the LHS of (4.1.2) reads

$$-\frac{\lambda f}{q_0 k_0}(1+kt_1)\exp\left(\frac{q\frac{N}{n}(t_1-t_c)}{1+\exp(-r_1)}-r_1\right)\left[\frac{q\frac{N}{n}(t_1-t_c)}{1+\exp(-r_1)}-1\right]-\frac{aq\frac{N}{n}\exp(-r_1)}{(1+\exp(-r_1))^2}, \tag{4.1.3}$$

while the t_1-derivative reads

$$-\frac{\lambda f}{q_0 k_0}(1+\exp(-r_1))\exp\left(\frac{q\frac{N}{n}(t_1-t_c)}{1+\exp(-r_1)}\right)\left[k+\frac{q\frac{N}{n}(1+kt_1)}{1+\exp(-r_1)}\right]-\frac{k^2}{\left(1+kt_1\right)^2}. \tag{4.1.4}$$

Evaluated at $(r_1, t_1) = (r_1^*, t_1^*)$ the slope of the line tangent to the equilibrium curve is given by

$$\frac{\frac{\lambda f}{q_0 k_0}(1+kt_1)\exp\left(\frac{q\frac{N}{n}(t_1^*-t_c)}{1+\exp(-r_1^*)}-r_1^*\right)\left[\frac{q\frac{N}{n}(t_1^*-t_c)}{1+\exp(-r_1^*)}-1\right]+\frac{aq\frac{N}{n}\exp(-r_1^*)}{(1+\exp(-r_1^*))^2}}{\frac{\lambda f}{q_0 k_0}(1+\exp(-r_1^*))\exp\left(\frac{q\frac{N}{n}(t_1^*-t_c)}{1+\exp(-r_1^*)}\right)\left[k+\frac{q\frac{N}{n}(1+kt_1^*)}{1+\exp(-r_1^*)}\right]+\frac{k^2}{\left(1+kt_1^*\right)^2}}. \tag{4.1.5}$$

It is immediately clear that the denominator in (4.1.5) is always positive so that the monotonicity of the equilibrium curve can change only through changes in the sign of the numerator. This is unlike the example discussed in Scaramangas and Broom (2022) where sign changes were attributed to the denominator and manifest as vertices at which the line tangent were vertical. Here, we observe that if t_1^* is sufficiently low (this can be the case when λ is low) the term in square brackets can be made negative enough to make the numerator negative, such that t_1^* decreases as r_1^* increases. Likewise, when the background mortality rate λ is high enough the associated term in square brackets is positive so that the numerator (and fraction) is positive overall and the level of defence t_1^* increases with increasing levels of conspicuousness. Changes in monotonicity are observed in Figures 4.6/7 at the end of section 4.4 (following forward 20 pages) and discussed therein.

From (4.1.4) it is clear that the terms on the LHS of (4.1.2) are decreasing with respect to t_1^*. Likewise, it is observed from (4.1.3) that when t_1^* is sufficiently low/high (e.g. λ is high/low) the first term in (4.1.2) is increasing/decreasing with respect to r_1^* while the fourth term is monotonically increasing with respect to r_1^* (independent of t_1^*). Suppose that (r_1^*, t_1^*) satisfies the equilibrium condition (4.1.2) for some low enough value of λ that the overall sign of (4.1.3) is negative. In this case, a marginal increase in r_1^* will (by assumption) reduce the LHS of (4.1.2) which, on account of (4.1.4) being negative, must be compensated by a reduction in t_1^*. The latter suggests that when λ is sufficiently low the equilibrium level of defence (defined implicitly through (4.1.2)) is decreasing with respect to conspicuousness. Likewise, we can assume

that (r_1^*, t_1^*) satisfies (4.1.2) for some value of λ that is sufficiently high that (4.1.3) is positive. In such a situation increasing r_1^* would cause the LHS of (4.1.2) to increase so that to restore equilibrium this must be compensated with an increase in t_1^*, suggesting that for high enough λ the equilibrium defence increases with conspicuousness.

We remark that if $a = 0$ - this relates to cases (i) and (iii) examined below - the numerator of (4.1.5) is zeroed when

$$t_1 = t_c + \frac{1 + \exp(-r_1)}{q\frac{N}{n}} =: c(r_1). \tag{4.1.6}$$

Suppose that (r_1^*, t_1^*) satisfies (4.1.2). If $t_1^* < c(r_1^*)$ the slope of the line tangent to the curve (4.1.2) at (r_1^*, t_1^*) is positive; if $t_1^* = c(r_1^*)$ the slope of the line tangent to the curve (4.1.2) at $(r_1^*, c(r_1^*))$ is zero and finally, if $t_1^* > c(r_1^*)$ it follows that the slope of the line tangent at (r_1^*, t_1^*) is negative. As discussed more extensively in the previous chapter as well as inScaramangas and Broom (2022) and contrary to what prevailing theory contends the relationship of aposematic traits may not need not be an increasing one. As for case (iv) there is (to our knowledge) no way of determining the resident fitness at equilibrium and one may resort to numerical methods to achieve this.

In case (i) it is immediately clear that setting $\lambda = 0$ and $a = 0$ into (4.1.2) eliminates the first and fourth terms on the LHS so that prey defence (at ESS) is not associated with the conspicuousness. That is

$$t_1^* = \frac{1}{f} - \frac{1}{k}, \tag{4.1.7}$$

for all $r_1^* \geq 0$. This suggests that mutants with incrementally higher levels of defence (compared with the residents) pay a price for reproducing at a slower rate, but are better defended against attacks that are potentially lethal so that at the (unique) ESS level of defence the two components balance as in (4.1.7). An important assumption of the model (see chapter 2) is that investment in defences (but not in bright colourations) is costly and this is reflected in the negative dependence on t of the fecundity function $F(t)$. Indeed, once the level of toxicity described in (4.1.7) is reached, resident strategies with different signalling component may have different overall levels of fitness, but cannot be invaded by mutants that are (incrementally) more/less defended (since the trade-off between F and K is exact). An alternative (but more equation-intensive) approach would be to impose that investment in bright colourations also impacts the fecundity negatively. Doing so would introduce a dependence on the conspicuousness of the ESS level of defence, even within the regime described by (i).

In case (ii) the level of defence satisfying (4.1.2) can be provided explicitly in terms of the conspicuousness as

$$t_1^*(r_1^*) = \frac{1}{f - \dfrac{aq\frac{N}{n}}{1 + \exp(-r_1^*)}} - \frac{1}{k} \tag{4.1.8}$$

for all $r_1^* \geq 0$. The latter suggests that the ESS level of defence is increasing with increasing levels of the conspicuousness and that the increase is sharper for larger values of the parameter a; we direct the reader to Broom et al., (2008) for a more careful consideration of this example. The situation in (ii) is different to (i) in that mutation is now assumed to occur in clusters of size a, whose size influences their perceived aversiveness and the probability that predators visiting their site mount attacks on them. So while it is true that for mutants with incrementally larger levels of defence the cost to fecundity must be counterbalanced by the benefit of escaping potentially lethal attacks, there is in (ii) the effect of additional protection against predation accrued by the presence of better-defended mutants in a group that is sizeable. The relationship

between conspicuousness and defence is more sharply increasing when the associated level of defence is smaller (see Fig. 4.3a) since prey must broadcast their aversiveness more strongly to reduce predation. Beyond a certain level of defence further increases in the conspicuousness have diminishing returns on the rate that they are attacked.

We should add that if (r_1^*, t_1^*) satisfies (4.1.2) then strategy (r_1^*, t_1) with $t_1 < t_1^*$, which includes the origin, fails (2.3.92)/(2.3.95). This is attributed to the fact that the LHS of (4.1.2) decreases with respect to positive changes in the argument t_1 and since (by assumption) the LHS is zero for $t_1 = t_1^*$ and $r_1 = r_1^*$ it follows that the LHS is positive for values $t_1 < t_1^*$. The argument could be repeated for choices of $t_1 > t_1^*$ in which case the LHS of (4.1.2) would be negative by the property that the mutant fitness in the vicinity of the resident value is almost everywhere $\mathcal{C}^l$ with $l \geq 2$. The interpretation in either case suggests that levels of with $t_1 < t_1^*$ are at risk of invasion against mutants that are more toxic - LHS of (4.1.2) is positive - while levels of defence with $t_1 > t_1^*$ risk being invaded by less toxic mutants - LHS of (4.1.2) is negative. Through inspection we have deduced that strategies (r_1^*, t_1^*) satisfying (4.1.2) uninvadable along t (since levels of defence below t_1^* are invaded by the more toxic types and levels beyond t_1^* are invaded by the less toxic types). We make this claim more formal in the lines that follow.

In Broom et al. (2008) it was shown that such strategies are satisfy (2.3.92)/(2.3.95) for the case $\lambda = 0$. We extend the substitution method found therein in a straightforward manner to establish that it holds for all values of $\lambda \geq 0$. We proceed by considering the cases $t_1^* > 0$ and $t_1^* = 0$ separately.

Consider strategy (r_1^*, t_1^*) satisfying (4.1.2). Substitution of (4.1.1) into (2.3.95) now amounts to

$$-\frac{\lambda f^2}{q_0 k_0}(1+\exp(-r_1^*))(1+kt_1^*)\exp\left(q\frac{N}{n}\frac{t_1^*-t_c}{1+\exp\left(-r_1^*\right)}\right) - f^2 + \frac{2k^2}{(1+kt_1^*)^2}$$

$$+\frac{2aq\frac{N}{n}}{1+\exp(-r_1^*)}\frac{k}{1+kt_1^*} + \frac{a^2q^2\frac{N^2}{n^2}}{(1+\exp(-r_1^*))^2} > 0. \tag{4.1.9}$$

We set

$$\Lambda := \frac{\lambda f}{q_0 k_0}(1+kt_1^*)(1+\exp(-r_1^*))\exp\left(q\frac{N}{n}\frac{t_1^*-t_c}{1+\exp(-r_1^*)}\right) \tag{4.1.10}$$

and re-arrange (4.1.2) so that

$$\frac{aq\frac{N}{n}}{1+\exp(-r_1^*)} = \Lambda + f - \frac{k}{1+kt_1^*}. \tag{4.1.11}$$

Condition (4.1.9) now amounts to

$$-\Lambda f - f^2 + \frac{2k^2}{(1+kt_1^*)^2} + \frac{2k}{1+kt_1^*}\left(\Lambda + f - \frac{k}{1+kt_1^*}\right) + \left(\Lambda + f - \frac{k}{1+kt_1^*}\right)^2 > 0 \tag{4.1.12}$$

and simplifies to the trivial inequality

$$\Lambda^2 + \Lambda f + \frac{k^2}{(1+kt_1^*)^2} > 0. \tag{4.1.13}$$

We have therefore demonstrated that for all values of the parameter $\lambda \geq 0$ strategies (r_1^*, t_1^*) on the curve given through (4.1.2) satisfy (4.1.9).

Strategies with $t_1^* = 0$ must satisfy (2.3.88) if $r_1^* > 0$, which is the same as equality in (4.1.2) is replaced with inequality < 0. Since from Broom et al. (2006) it is known that strategies of the form $\{(\rho, \tau) : \rho > 0\ \tau = 0\}$ fail (2.3.90) it follows that the origin $\{(0,0)\}$ is the only possibility for a non-toxic strategy to

be ESS. For strategy $(r_1^*, t_1^*) = (0, 0)$ to be a local ESS it must satisfy (2.3.87), which considering (4.1.1) amounts to

$$-\frac{2\lambda f}{q_0 k_0}\exp\left(-q\frac{N}{2n}t_c\right) - f + k + aq\frac{N}{2n} < 0 \tag{4.1.14}$$

and it is clear that there is sufficient freedom on the parameters to either satisfy or fail to satisfy the above inequality.

From looking at (2.3.92)/(2.3.96) and (2.3.93)/(2.3.97) it is clear that these conditions are unaffected by the rate of background mortality λ. From a practical standpoint, this is the case because it is convenient for purposes of stability to consider the normalised gradient of the mutant fitness, which factors this dependence out. Going beyond this, we observe that differences in the mutant fitness (along r) are associated with differences in the average life-span of prey through influencing the rates of predator detection, recollection and perceived aversiveness (by comparison with the resident appearance); it should be remarked that none of the above are affected by whether the threat of predation is large (i.e. by the value of λ) compared with threats outside of predation. Indeed, a given regime of background mortality applies to both the resident and the mutant and since incremental changes in the fitness of the latter (along r) are unaffected by the value of λ, the prospect of invasion by the latter is also unaffected by the value of λ.

While invasion along r does not depend on the parameter λ it does depend on the local clustering parameter a. This too comes from direct observation of (2.3.96) and (2.3.97) and admits a sensible remark; larger groups tend to be better recollected by predators that experience their type and further, the larger a group whose appearance deviates from (say, an aversive) resident majority the larger the fitness cost incurred to the mentioned group collectively.

A resident strategy with $r_1 > 0$ is stable in the r-direction if (2.3.96) and (2.3.97) both hold, which on account of (4.1.1) read

$$-\overleftarrow{\partial}_r P(r_1, t_1) \approx \exp(-r_1) - q\frac{N}{n}(t_1 - t_c)\left[\frac{a}{1+\exp(r_1)} + (1-a)v\right] < 0 \tag{4.1.15}$$

and

$$\overrightarrow{\partial}_r P(r_1, t_1) \approx -\exp(-r_1) + q\frac{N}{n}(t_1 - t_c)\left[\frac{a}{1+\exp(r_1)} - (1-a)v\right] < 0. \tag{4.1.16}$$

The normalised gradient of the mutant fitness along r (corresponding to quantities $-\overleftarrow{\partial}_r P$ and $\overrightarrow{\partial}_r P$ defined above) are referred to collectively as the *invasion fitness gradient* throughout this chapter. Cryptic strategies are stable in r if (4.1.16) holds with $r_1 = 0$. The $\approx$ notation is used to remind readers that the quantities on the LHS of the inequalities are not equal to the derivatives $-\overleftarrow{\partial}_r P$ and $\overrightarrow{\partial}_r P$ but have been scaled by $(\lambda + DKQ)^2/FDKQ$. For purposes of notational convenience - specifically in (4.1.18) and (4.1.17) - we treat these as equal. The difference in fitness (compared with the resident) of a mutant with incrementally smaller conspicuousness is to first order given as $-\overleftarrow{\partial}_r P dr$. We should mention that the quantities on the RHS of (4.1.15) and (4.1.16) exhibit horizontal asymptotes as $r_1 \to \infty$ which can be attributed to the choice of function for $D(r)$ in (4.1.1), which plateaus in this limit.

A linear aversiveness function $H(t)$ as in (4.1.1) is both a technically sensible and a biologically plausible choice. As a consequence the LHS of the inequalities in (4.1.15) and (4.1.16) are linear in t_1, which allows us to express explicitly the toxicity in terms of the conspicuousness and $-\overleftarrow{\partial}_r P$ or $\overrightarrow{\partial}_r P$. We have the useful

substitutions

$$t_1 = \frac{\left[-(-\overleftarrow{\partial}_r P) + \exp(-r_1)\right](1+\exp(r_1))}{q\frac{N}{n}[a + v \times (1-a)(1+\exp(r_1))]} + t_c =: g_-\big(r_1, -\overleftarrow{\partial}_r P\big) \tag{4.1.17}$$

and

$$t_1 = \frac{\left(\overrightarrow{\partial}_r P + \exp(-r_1)\right)(1+\exp(r_1))}{q\frac{N}{n}[a - v \times (1-a)(1+\exp(r_1))]} + t_c =: g_+\big(r_1, \overrightarrow{\partial}_r P\big), \tag{4.1.18}$$

which we can utilise in (4.1.2) to obtain implicit expressions for the invasion gradient of the mutant along r $-\overleftarrow{\partial}_r P(r_1^*, t_1^*)$ and $\overrightarrow{\partial}_r P(r_1^*, t_1^*)$ at equilibrium. These are

$$\frac{\lambda f}{q_0 k_0}(1+kg_\mp)(1+\exp(-r_1^*))\exp\left(q\frac{N}{n}\frac{g_\mp - t_c}{1+\exp(-r_1^*)}\right) + f - \frac{k}{1+kg_\mp} - \frac{aq\frac{N}{n}}{1+\exp(-r_1^*)} = 0, \tag{4.1.19}$$

which we represent as orange and cyan curves in Figures 4.2c, 4.4c and 4.5c. The substitution method outlined above is general and especially useful in cases where the level of defence t_1^* in (4.1.2) cannot be expressed explicitly in terms of the conspicuousness (such as when $\lambda > 0$). However, in cases where $\lambda = 0$ we observe that $-\overleftarrow{\partial}_r P^*$ and $\overrightarrow{\partial}_r P^*$ can be evaluated directly by setting $t_1 = t_1^*$ in the RHSs of (4.1.15) and (4.1.16), making the above method superfluous.

The parameter v, which was introduced in (4.1.1) and which is present in (4.1.15) and (4.1.16) above can be understood as the predator's perception of small differences in the visual appearances of prey. We could for all intents and purposes think of this as the time a predator spends investigating a prey animal before deciding to mount an attack. The larger this quantity is the better the predators are at telling apart small differences in the conspicuousness of warning signals (they spend less time investigating it); the smaller this is the worse they are. As we detail, the significance of this term is different for attractive prey with $t_1 < t_c$ than it is for aversive prey with $t_1 > t_c$.

If $t_1 < t_c$ it is easy to observe that (4.1.15) cannot be solved for any sensible choice of v. This result is in line with the more general reasoning of Broom et al. (2006), in which it is argued that the conspicuous signalling of strategies that are non-aversive - i.e. drawn from $\{\rho > 0;\ 0 \leq \tau \leq t_c\}$ - risk invasion from mutations with incrementally smaller signalling component through failing. The same is not true for (4.1.16) however, which can be solved for values of v below the threshold on the RHS of

$$v < \frac{\exp(-r_1) + aq\frac{N}{n}\frac{|t_1 - t_c|}{1+\exp(-r_1)}}{(1-a)q\frac{N}{n}(t_1 - t_c)}. \tag{4.1.20}$$

The direction of this inequality demonstrates that for an attractive resident strategy (cryptic) to successfully resist invasion of a more conspicuous mutant, the predator cannot be exceptionally observant, otherwise it would avoid attacking the mutant altogether making the latter comparatively fitter.

Residents playing aversive strategies with $t_1 > t_c$ can resist invasion by less/more conspicuous mutants

if (4.1.15) and (4.1.16) can be solved for values of v large enough that

$$v > \frac{\left|\exp(-r_1) - aq\frac{N}{n}\frac{t_1 - t_c}{1+\exp(-r_1)}\right|}{(1-a)q\frac{N}{n}(t_1 - t_c)}. \tag{4.1.21}$$

The direction of the inequality is also justified in this instance; we would expect an aversive majority of residents to withstand invasion provided the predator is sufficiently observant to detect incremental differences in conspicuousness. Mutants that look different to a majority of prey that is perceived as aversive pay a price for this and the cost of that decision is magnified by the predator's ability to perceive such differences.

Absolute resident fitness

We now discuss resident fitness, which was introduced in (2.3.4). Considering the example function in (4.1.1) this amounts to

$$P_1(r_1, t_1) = \frac{f_0 \exp(-ft_1)}{\lambda + \dfrac{q_0 k_0}{(1+\exp(-r_1))(1+kt_1)\exp\left(q\frac{N}{n}\dfrac{t_1 - t_c}{1+\exp(-r_1)}\right)}}. \tag{4.1.22}$$

For the purposes of understanding the outcomes of the simulations in the next section it is of interest to determine how this quantity varies when the resident traits are drawn from the curve (4.1.2). There are four cases to consider: (i) $\lambda = 0, a = 0$; (ii) $\lambda = 0, a > 0$; (iii) $\lambda > 0, a = 0$ and (iv) $\lambda > 0, a > 0$.

The method for (i) and (ii) involves solving for t_1^* explicitly in terms of r_1^* in (4.1.2) - observe from (4.1.7) and (4.1.8) that this is possible - and replacing t_1 in (4.1.22) with the equilibrium value t_1^*. For (i) the equilibrium condition we set $t_1^* = 1/f - 1/k$ in (4.1.22) to obtain the required result. Likewise for (ii) we set (4.1.8) into (4.1.22).

For (iii) we proceed by re-arranging (4.1.22) so that

$$\frac{q_0 k_0}{(1+\exp(-r_1))(1+kt_1)\exp\left(q\dfrac{N}{n}\dfrac{t_1 - t_c}{1+\exp(-r_1)}\right)} = \frac{f_0\exp(-ft_1)}{P_1} - \lambda. \tag{4.1.23}$$

Substitution of this term into the equilibrium leads to expression

$$-\frac{\lambda f P_1^*}{f_0\exp(-ft_1^*) - \lambda P_1^*} - f + \frac{k}{1+kt_1^*} = 0, \tag{4.1.24}$$

which is equivalent to

$$t_1^* = \frac{1}{f} - \frac{1}{k} - \frac{\lambda P_1^*}{f f_0}\exp(ft_1^*). \tag{4.1.25}$$

The latter can be solved in terms of the principal branch of the *Lambert W-function* (this is such that $W_0(x)\exp(W_0(x)) = x$ provided $x \geq 0$ - the more mathematically-minded reader is encouraged to consult Corless et al., 1996 for an in-depth discussion of the properties and applications of this function). Using a

known ansatz we arrive at an explicit expression for the level of defence t_1^* in terms of the resident fitness

$$t_1^* = \frac{1}{f} - \frac{1}{k} - \frac{1}{f} W_0 \left(\frac{\lambda P_1^*}{f_0} \exp\left(1 - f/k\right) \right) =: G(P_1^*). \tag{4.1.26}$$

We note that we have made use of the shorthand notation $P_1^* \leftrightarrow P_1(r_1^*, t_1^*)$ to denote the value of the resident fitness along the curve (4.1.2). Swapping t_1^* for $G(P_1^*)$ in equality (4.1.2) leads to an implicit expression for the resident fitness along this curve

$$\frac{\lambda f}{q_0 k_0}(1 + \exp(-r_1^*))(1 + kG(P_1^*)) \exp\left(q \frac{N}{n} \frac{G(P_1^*) - t_c}{1 + \exp\left(-r_1^*\right)} \right) + f - \frac{k}{1 + kG(P_1^*)} = 0. \tag{4.1.27}$$

4.2 A description of the simulation

Our simulations explicitly model all the individual members of a finite prey population. Individuals will potentially play different strategies, and the performance of individuals will depend on both their own strategy and the distribution of strategies of individuals that they interact with. A similar approach to addressing questions in the evolution of aposematism was taken by M. P. Speed and Ruxton (2005), and we further develop their approach. Here we represent evolution by selectively removing individuals from the population and replacing them with versions of other individuals. Prey phenotypes that perform well in the current population are more likely to contribute versions of themselves to the next generation. This mimics the effect of differential fitness in real populations, and is a common approach in evolutionary studies and beyond – often being labelled a genetic algorithm approach (G. D. Ruxton and Beauchamp, 2008). More generally, individual-based modelling is well established in the study of questions in evolutionary ecology (Zakharova et al., 2019).

The simulation assumes a population of N prey predated by n predators and playing strategies (r_i, t_i) with $i = 1, ..., N$ - to avoid confusion we restrict notations involving the iteration number only to where necessary (see the birth-death process detailed below). The specification of an individual's strategy directly determines the rate at which it reproduces (as $F_i = F(t_i) = f_0 \exp(-f t_i)$), the rate at which it is detected by predators ($D_i = D(r_i) = 1/(1 + \exp(-r_i))$) and the rate at which a mounted attack results in death ($K_i = K(t_i) = k_0/(1 + k t_i)$), as well as the aversiveness of the predator's experience (as $H_i = H(t_i) = t_i - t_c$) and the rate at which such experiences are recollected ($L_i = L(r_i) = 1/(1 + \exp(-r_i))$). The specification of such quantities over the population is realised using lists ($1 \times N$ vectors). In contrast, the perceived visual similarity of prey is stored in the $N \times N$ symmetric and unit-diagonal matrix $\mathcal{S}$ defined as

$$(\mathcal{S})_{ij} := S(r_i, r_j) = max\left(1 - v\left|r_i - r_j\right|, 0\right) \text{ for all } i, j = 1, ..., N. \tag{4.2.1}$$

The realisation of the i^{th} row of the matrix in (4.2.1) specifies the aversiveness $\mathcal{I}_i$ of that prey as perceived by the average predator through the rule

$$\mathcal{I}_i = \frac{1}{n} \left[(aN - 1) L(r_i) H(t_i) + \frac{(1-a)N}{N-1} \sum_{j=1, j \neq i}^{N} L(r_j) H(t_j) S(r_i, r_j) \right], \tag{4.2.2}$$

where the term

$$\frac{1}{N-1}\sum_{j=1,j\neq i}^{N} L(r_j)H(t_j)S(r_i,r_j) \tag{4.2.3}$$

in (4.2.2) is the aversivess of the average prey (excluding the focal individual i). Implicit in (4.2.2) is the assumption that (a) when encountering a prey and calculating its aversiveness $\mathcal{I}_i$, a predators weighs the prey individual it is currently facing as a proportion a of the entire population (independent of phenotype, mutant-status, or even population size), and this happens with every prey that is encountered by a predator in the simulations. The implementation of the local relatedness in the infinite population ESS analysis is different (a is evaluated as a proportion of individuals in the site). (b) Expression (4.2.2) represents an average (factor $1/n$) over the predator's experiences of prey and indeed an average over the prey that these encounter (factor $1/(N-1)$ excludes the focal individual - see related explanation in section 2). The probability Q_i that an attack is mounted on i depends on (4.2.2) through $Q(\mathcal{I}_i) = q_0\exp(-q\mathcal{I}_i)$ and its fitness is hence given as

$$P_i = P(r_i,t_i) = \frac{f_0\exp(-ft_i)}{\lambda + \dfrac{k_0 q_0}{(1+\exp(-r_i))(1+kt_i)\exp(q\mathcal{I}_i)}}. \tag{4.2.4}$$

We should remark that parameter a plays a role in the calculation of the fitness of individuals in the simulation (through Q), which in turn affects the likelihoods of reproduction – however it plays no part in the nature of that reproduction (i.e. in the number of offspring, or the effect of mutation).

The simulation tracks the evolution of traits for a number of distinct prey populations in the following manner. It commences at $m = 0$ where the index $m = 0, 1, 2, ..., M$ specifies the iteration number and can be understood as the number of *birth-death events* that have preceded the population in question (the details of this processes are provided below). After a fixed number of iterations has passed, which is determined by the *averaging frequency* g the population traits are averaged and the averaged pair of values is represented as a point in the strategy space of averages. A straight line segment (starting at the initial strategy) is drawn between consecutive points, such that the union of segments forms a *trajectory* for that population. The number of segments making up a population's trajectory is given as M/g. Trajectories of this type are drawn for populations playing a number of distinct starting strategies.

Prey populations succeed one another by means of a *birth-death process*, whose details are as follows. A small sample of pN prey is selected at random to reproduce and their offspring replace an equally-sized sample. We remark that prey may be selected to reproduce more than once (i.e. give birth to more than one offspring) and are thus considered with multiplicity on the list consisting of parents. It is also possible for the same individual to reproduce and to be replaced (by its own offspring) at the end of the same iteration. The probability that an individual is selected to reproduce l times after pN trials (with replacement) is a binomially-distributed random variable

$$\mathfrak{P}(i \text{ becomes parent } l \text{ times}) = \binom{pN}{l} W_i^l (1-W_i)^{pN-l}, \text{ for } l = 1, ..., pN \tag{4.2.5}$$

where W_i is a comparative measure of fitness defined as

$$W_i := \frac{P_i}{\sum_{j=1}^{N} P_j} \tag{4.2.6}$$

with P_i given as in (4.2.4). We should add that for large enough populations we expect the comparative fitness of any one individual to be relatively small and therefore the distribution in (4.2.5) to be approximately Poisson distributed with parameter pNW_i. A *generation* can be understood as the average number of iterations (birth-death events) required for all the individuals in a population to be replaced. We stress that while alternative interpretations of a generation are possible, from the point of view of the simulation a generation is synonymous with the average number of birth-death events required for the individuals comprising a certain population to be completely replaced.

Prey traits are subject to random mutation (in the sense that the offspring values can vary continuously within a small margin of error centred at the parent value) and this is encoded into the birth process. We remark that toxicity and conspicuousness are traits determined by common environmental factors (including predation threat and availability of food resources among others) and are likely polygenic, since few phenotypic traits have a single gene origin. Aposematic traits exhibit notable differences depending on the species in question (the genetic origin of traits could provide a possible explanation for this). Furthermore, the specific mode of interaction of one trait with the other is (to our knowledge) mostly unknown. It is therefore the natural option for purposes of simulation to assume that mutation in one trait does not influence mutation in the other (i.e. mutation in either trait is independent) versus a more specific (and controversial) assumption about their mode of interaction. In the same spirit, we remark that it is possible for mutation to occur in both traits during a single birth process. To be specific we say that if the offspring of individual i replaces individual j in transitioning from the m^{th} to the $m+1^{st}$ iteration, the probability that either trait is carried through to the offspring is given as 95%. We write

$$\mathfrak{P}\left(r_j^{(m+1)} = r_i^{(m)}\right) = \mathfrak{P}\left(t_j^{(m+1)} = t_i^{(m)}\right) = 0.95, \tag{4.2.7}$$

while the probability that any of the traits change is given as

$$\begin{aligned}\mathfrak{P}\left(r_j^{(m+1)} \in \left[r_i^{(m)} - \delta r, r_i^{(m)}\right) \sqcup \left(r_i^{(m)}, r_i^{(m)} + \delta r\right]\right) = \\ = \mathfrak{P}\left(t_j^{(m+1)} \in \left[t_i^{(m)} - \delta t, t_i^{(m)}\right) \sqcup \left(t_i^{(m)}, t_i^{(m)} + \delta t\right]\right) = 0.05.\end{aligned} \tag{4.2.8}$$

From context it should be clear that the mutation range during the described birth process is precisely the closed rectangle with dimensions $2\delta r \times 2\delta t$ centred at the parent value. As a consequence of independence in trait mutations we also remark that the probability that both parent traits are carried through to the offspring is $0.95^2 \approx 0.9025$, while the the probability that both traits change is $0.05^2 = 0.0025$. We should also remark that if a trait changes the step length is chosen uniformly from within the mutation range of the trait in question. For the first trait we write

$$\mathfrak{P}\left(r_j^{(m+1)} \in \delta x\right) = 0.05\frac{\delta x}{2\delta r} \tag{4.2.9}$$

to demonstrate the probability that if it increases (or decreases) its precise value is within the interval $\delta x \subset \left(r_i^{(m)}, r_i^{(m)} + \delta r\right]$ or within $\left[r_i^m - \delta r, r_i^{(m)}\right)$ if it decreases.

4.3 Solutions with vanishing background mortality

In this section we consider the simplest scenario in which $\lambda = 0$ and treat cases (i) $a = 0$ and (ii) $a > 0$ separately. We make use of the theory developed earlier and focus our attention on (a) the predicted form of ESS, (b) the resident fitness at equilibrium and (c) the invasion fitness gradient (along r). This style of presentation exposes the reader to gradually increasing levels of complexity and is also adopted in the section following this, which deals with the cases (iii) and (iv) in which $\lambda > 0$. We should also remind the reader that the theoretical/predictions component of this section is based on the existing works of Broom et al. (2006) and Broom et al., (2008).

The $a \to 0$ limit

The plots of Figure 4.2 showcase our findings for case (i) in which $\lambda = 0$ and $a = 0$.

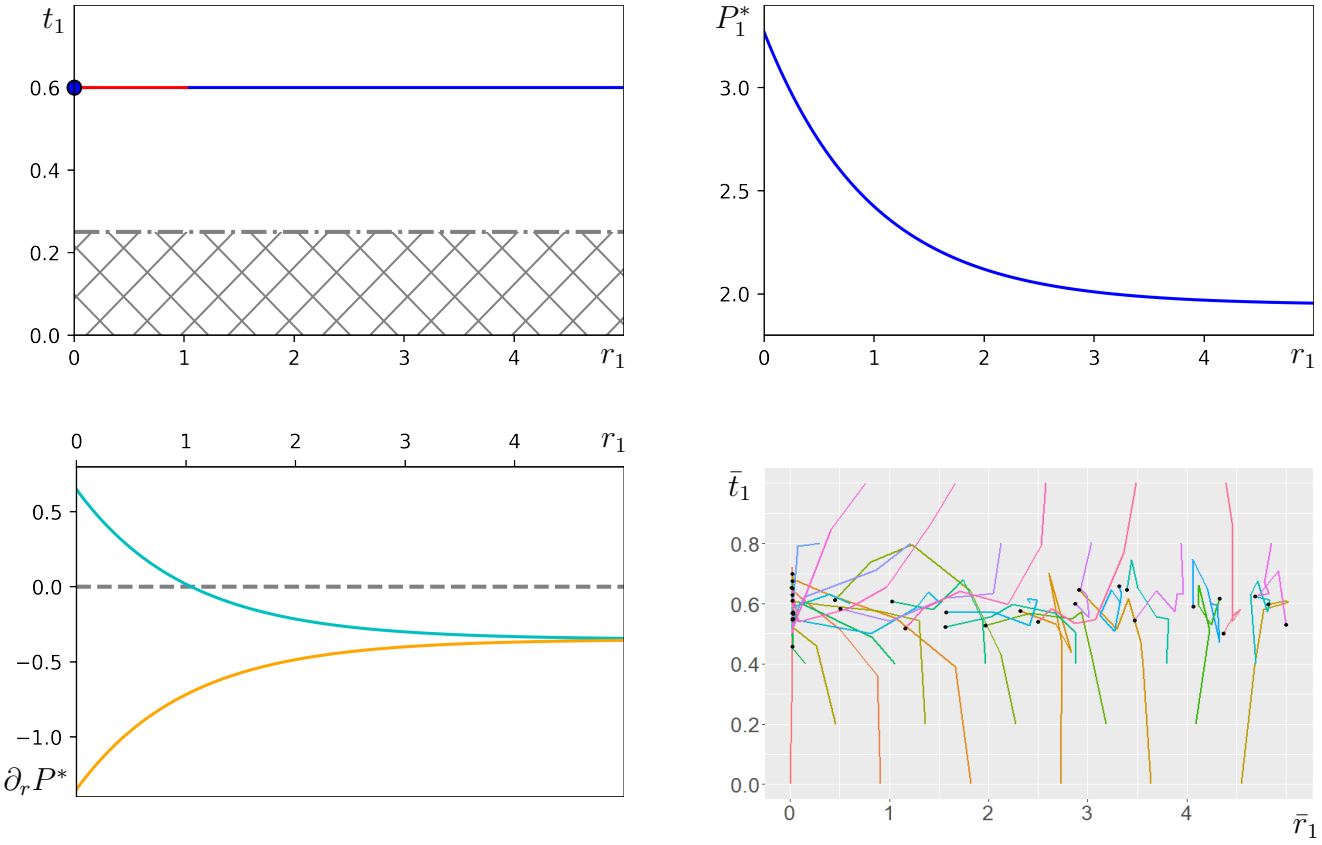

Figure 4.2: Parameter values $\lambda = 0, a = 0, f_0 = q_0 = k_0 = 1, f = 1, k = 2.5, t_c = 0.25, q = 0.1, N = 100, n = 10$ and $v = 1$ **(a)** [Top left] Strategies within the grey-shaded region $\{(r_1, t_1) : r_1 > 0, t_1 \leq 0.25\}$ fail (4.1.15). Unique cryptic ESS at $(r_1^*, t_1^*) = (0, 0.6)$ shown with blue marker, succeeded by strategies on (4.1.7) that fail (4.1.15) for $r_1^* < 1.05$ but are stable beyond that value (blue section). **(b)** [Top right] Resident fitness evaluated along (4.1.7) as per (4.1.22). **(c)** [Bottom left] Invasion fitness gradient along r - see (4.1.19) as well as (4.1.15) and (4.1.16) - for incrementally less (cyan curve) and incrementally more conspicuous mutants (orange curve). **(d)** [Bottom right] Average population traits plotted as trajectories with averaging frequency $g = 2,000$. Black markers represent the average traits of a single population after $10,000$ iterations.

The black markers in Figures 4.2(d) and 4.3(d) indicate that the majority of prey populations eventually

converge close to the predicted equilibrium toxicity level of $1/f - 1/k$ as given in (4.1.7). We also remark that the lower the initial conspicuousness of the population the stronger the component of its associated trajectory toward crypsis. In Figure 4.2d this would be expected for initial conspicuousness values below the cut-off specified through (4.1.15) but we observe that even populations starting from evolutionarily stable strategies are invaded by less conspicuous types. In observing the plots of Figure 4.2 alone one could speculate that this is attributed to the resident fitness being higher at crypsis. However, from Figure 4.3 (below) we deduce that this is unlikely the case, since there populations evolve against increasing resident fitness and toward crypsis where the less conspicuous mutants are increasingly advantageous. Presently, we make a number of important remarks about the invasion fitness gradient, which we use throughout to interpret the results of simulations. Following this, we discuss resident fitness and compare its impact on the evolution of prey traits alongside invasion fitness.

As it happens, plots of the invasion fitness gradient (orange and cyan curves in Figures 4.2c, 4.3c as well as in 4.4c and 4.5c) are consistent with the mutant landscape in the vicinity of the resident value along r. That is, when the cyan/orange curve is above the r-axis a resident population with that level of conspicuousness is predicted to be invasible by less/more conspicuous types (see cyan curves for $\bar{r}_1 < 1$ in Figures 2c as well as for $\bar{r}_1 < 0.5$ in 4c and the observed pull toward crypsis in 4.2d and 4.4d). In the majority of the cases we explore both the cyan and orange curves sit below the $r = 0$ axis (infinite population ESS analysis would deem such cases as evolutionarily stable along r) and the height below which they do so indicates how "worse-off" mutation in that direction is. An interesting effect of finiteness of the prey population is that strategies predicted as being evolutionarily stable (along r) still have a chance of being invaded if alternative mutations are not too worse-off. For instance, strategies with initial conspicuousness $1 < \bar{r}_1 < 3$ in Figure 2(d) and $\bar{r}_1 < 3$ in Figure 4.3 that are still invaded (through chance) by less conspicuous mutants.

For high enough levels of conspicuousness it is observed that the orange and cyan curves in Figures 4.2(c) and 4.3(c) converge (horizontally) to a common value. Technically, this can be attributed to our chosen forms for D (and therefore L - see (4.1.1)), which exhibit a plateau for high enough levels of conspicuousness (already bright signals do not impact detection/recollection further). In such situations mutation in either direction leads to equally bad outcomes suggesting there is no directional selection associated with the invasion fitness gradient along r. The further below the r-axis the asymptote is reached the more worse-off mutations are predicted to be so that not only is invasion equally likely in either direction, the probability of this occurring shrinks. Indeed, from a quick reading at $\bar{r}_1 > 3$ it is clear that the trajectories in Figure 4.3(d) appear less incidental than in Figure 4.2(d), where the associated asymptote is above -0.5 (compared with -3 in Figure 4.3c). We conclude that the smaller the distance between the cyan and orange curves the smaller the difference in selection between either direction and the smaller the value that these converge to the more unlikely invasion (in either direction) is overall.

While the simulations in Figures 4.2 and 4.3 fall under the same regime with respect to local clustering and background mortality ($a = 0$ and $\lambda = 0$) these show two principal differences, whose impact we explore further. The first difference is with respect to the invasion fitness gradient: In Figure 4.2, overall selection for smaller conspicuousness is strongest (and manifest as a stronger pull toward crypsis) and when selection is absent (high $\bar{r}_1$) randomness (seen in the time evolution of trajectories) is higher because invasion is likelier (though equally so in either direction). In addition, we have concluded that identifying strategies as "stable" or "unstable" is of limited use when studying prey populations that are finite, unless these are complemented with more precise statements describing "how stable"/"how unstable" those strategies are.

The second difference is with respect to the resident fitness at equilibrium. Viewing Figure 4.2 it is

difficult to set aside the impact of absolute resident fitness because this is highest for low $\bar{r}_1$ where (mutant fitness led) selection for less conspicuous types is also strongest. However, we observe that reversing the direction of increase of absolute resident fitness (Figure 3) does not significantly affect the outcome of the simulations. For sufficiently high values of $\bar{r}_1$ (where directional selection associated with the invasion fitness gradient is low) we could have expected prey trajectories in Figure 4.3(d) to evolve in the direction of increasing conspicuousness. Instead, these appear to evolve in mostly a random fashion and we conclude that this measure of fitness has little effect on the evolution of prey traits. This could be because under low local relatedness ($a = 0$), resident fitness does not predict mutant fitness (which is the quantity determining the direction of evolution).

In particular, through the examples in Figures 4.2 and 4.3 we establish three important facts relating to the evolution of traits in finite prey populations: (i) ESS analysis provides accurate insight into the behaviour of finite populations even though notions of stability are not completely deterministic. (ii) Mutant fitness along r appears to be the stronger driver of changes in prey traits compared to the resident fitness. In fact, the probability of invasion along the r-direction depends continuously on how worse-off the mutant type is compared with the resident, as opposed to some absolute rule describing stability. (iii) In the absence of directional selection associated with incremental increases in mutant fitness (along r) and/or absolute resident fitness the evolution of traits is mostly random.

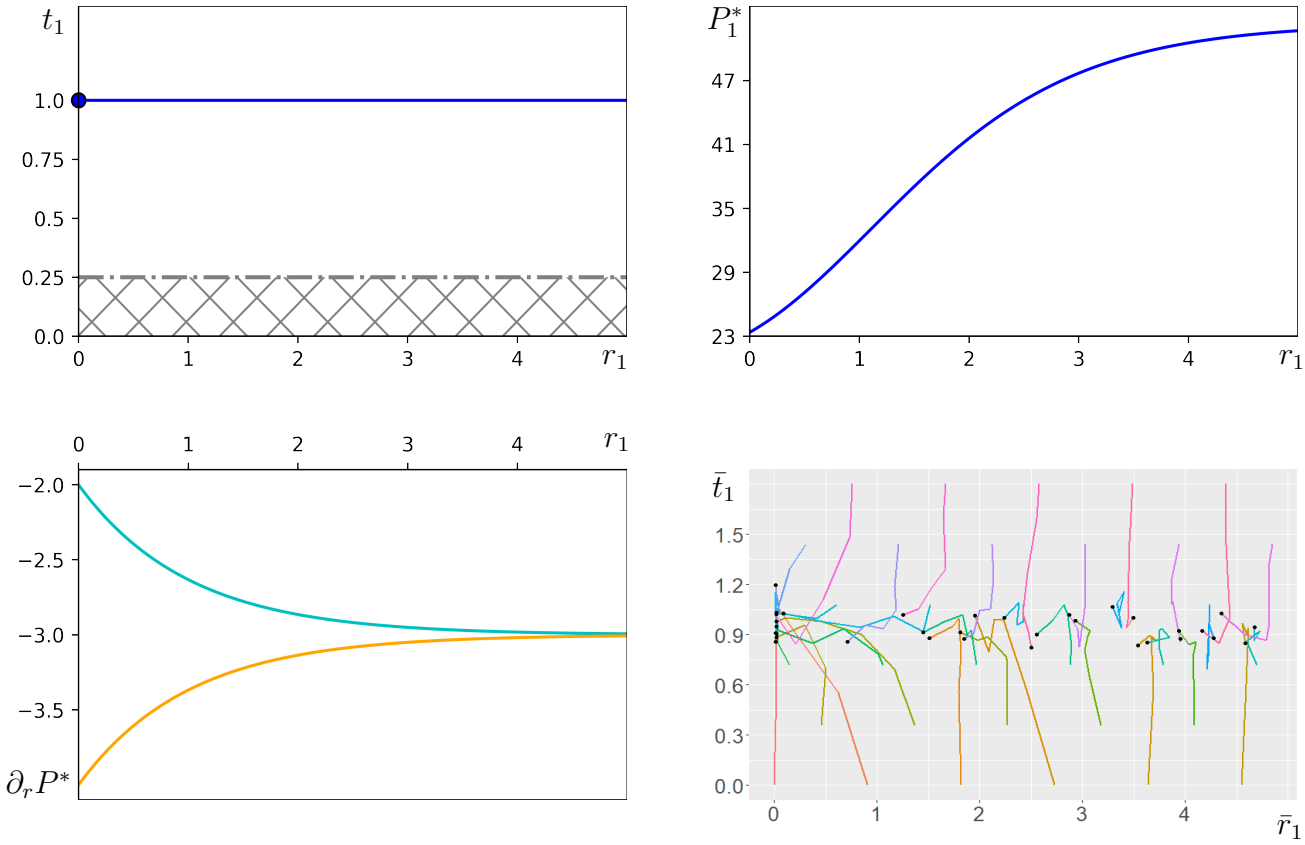

Figure 4.3: Parameter values $\lambda = 0, a = 0, f_0 = q_0 = k_0 = 1, f = 5/6, k = 5, t_c = 0.25, q = 0.4, N = 100, n = 10$ and $v = 1$ **(a)** [Top left] Grey-scale region $\{(r_1, t_1) : r_1 > 0, t_1 \leq 0.25\}$ contains strategies that fail (4.1.15). Solid blue marker at $(0, 1)$ is the unique cryptic ESS followed by a horizontal blue line of conspicuous ESSs evaluated at (4.1.7). **(b)** [Top right] Plot of resident fitness along (4.1.7) as per (4.1.22). **(c)** [Bottom left]: Invasion fitness gradient along r evaluated at equilibrium for incrementally less (cyan) and incrementally more (orange) conspicuous mutants as in (4.1.15) and (4.1.16) **(d)** [Bottom right]: Average population traits plotted as trajectories with averaging frequency $g = 2,000$. Black marks represent the average traits of a population after $10,000$ iterations.

Incorporating $a > 0$

The black markers in Figure 4.4(d) suggest that for most populations the average level of toxicity converges to the predicted equilibrium provided in (4.1.8). Populations starting with low conspicuousness risk being invaded by less conspicuous mutant types (the cyan curve in Figure 4.4(c) sits above the r-axis for $\bar{r}_1 < 0.5$) and the associated trajectories quickly converge to crypsis, as expected. For increasing levels of initial conspicuousness the leftwards component of the trajectories diminishes (more drastically than with $a = 0$ in Figures 4.2d and 4.3d) until it changes direction. This change in direction is recorded at $\bar{r}_1 \approx 1.5$, beyond which directional selection associated with the invasion fitness gradient vanishes (the cyan and orange curves in Figure 4.4c converge) while the absolute resident fitness continues to increase. A considerable proportion of the trajectories with initial conspicuousness $\bar{r}_1 > 2$ in Figure 4.4(d) are observed to evolve toward higher conspicuousness and we speculate that this can be traced back to the resident fitness.

Comparing these observations with those in Figures 4.2 and 4.3 we deduce that the impact of absolute resident fitness is more substantial when the size of the local relatedness parameter is greater. Indeed, in Figure 4.4(d) we observe evolution towards higher levels of resident fitness, especially in regions where there

is no directional selection associated with mutant fitness and in which mutants that are less conspicuous are notably worse-off compared with the residents. Crypsis tends to be the default and preferred strategy for a multitude of chemically defended prey and it is of interest to determine how and why aposematic solutions with a strong signalling component could instead admit a more viable option.

A plausible explanation for the above results can be found by considering positive frequency-dependent selection. Consider a mutant invading a resident population whose r strategy is similar but distinct. For our model there is a continuum of r values that are stable against invaders playing different r (both smaller and larger; this is because there is an inherent disadvantage for looking different from everybody else). For a pair of such strategies, A and B, an A population is stable against B invaders and a B population is stable against A invaders. Mutants can appear with higher or lower r values, and there will be a small probability of successful invasion, which is amplified by the size of the local relatedness parameter a. If this parameter is large enough then due to positive frequency dependence on initial invasion and the finiteness of the population, invaders can quickly reach a sufficiently high overall frequency through a sequence of drift related invasions. It is likely that once a certain (threshold) frequency is reached selection turns positive for the mutant (as there is now an inherent disadvantage to the residents for looking unlike the invading mutant group) leading such an invader to go to fixation.

As we observe in the simulations of Figure 4.4 the type with the higher resident fitness generally has a higher probability to invade the type with lower resident fitness than for the reverse invasion. Thus a sequence of drift related invasions of the kind discussed will tend to move the population in the direction of higher resident fitness. The higher the value of parameter a the greater the local frequency of the mutant at the start, and so the lower the advantage to the resident. This increases the probability of any invasion in either direction, but the increase is more marked in the direction of higher resident fitness because of its relative stability, so that increasing a amplifies the above effect.

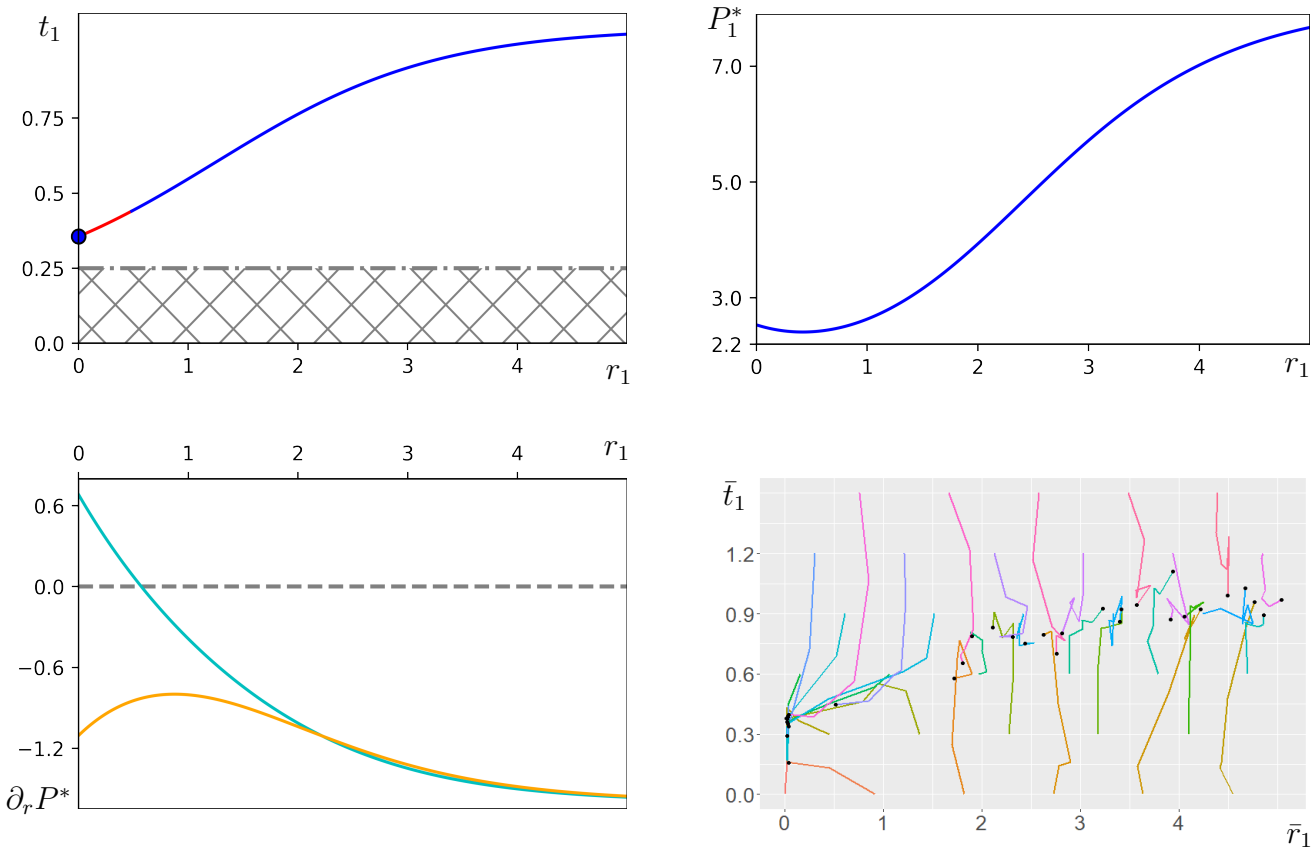

Figure 4.4: Parameter values $\lambda = 0, a = 0.5, f_0 = k_0 = q_0 = 1, f = 2.8, k = 5, t_c = 0.25, q = 0.4, N = 100, n = 10, v = 1$ **(a)** [Top left] Strategies inside grey-scaled region $\{r_1 > 0, t_1 \leq 0.25\}$ violate (4.1.15). Blue marker at $(0, 0.356)$ represents the unique cryptic solution, which co-exists alongside a continuum of conspicuous unstable (red) and stable ESSs (blue) on the curve of (4.1.8). **(b)** [Top right] Resident fitness as a function of conspicuousness along (4.1.8) conspicuousness. **(c)** [Bottom left] Cyan and orange curves represent the invasion fitness gradient given through (4.1.15) and (4.1.16) and evaluated along (4.1.8) **(d)** [Bottom right] Average population traits plotted as trajectories with averaging frequency $g = 2,000$. Black markers represent traits averaged over the population after $10,000$ iterations and mostly converge to $t_1^*(r_1^*)$ in (4.1.8) in Figure 4.4(a).

4.4 Solutions with non-zero background mortality $\lambda > 0$

In the previous subsection we established a set of empirical rules that can serve as a guide in our understanding of how aposematic traits evolve in finite prey populations that are subject to random mutation. We observed that while in the t-direction populations mostly evolve toward the predicted value, evolution along the r-axis is less straightforward. That is, one first has to consider whether there is directional selection for less or more conspicuous mutant types (for less conspicuous resident populations it tends to be the former) and particularly how much better/worse the type in question is compared with the resident. Second to this, we gauge the size of the directional selection (at equilibrium) with regards to the absolute resident fitness. It appears that this secondary cause can select for conspicuous solutions provided the local relatedness parameter is large enough and invasion in either direction is unlikely from the mutant fitness perspective.

In the present section we introduce non-zero rates of background mortality, initially in absence of local relatedness effects and finally including these. The presentation in this part places stronger emphasis on the outcomes of numerical simulation so as to showcase a larger breadth of examples within this less-explored

regime and more effectively observe the impact of varying the background mortality rate on finite populations.

The $a \to 0$ limit

In the plots of Figures 4.5 and 4.6 we showcase our predictions and findings for the case (iii) in which $\lambda > 0$ and $a = 0$.

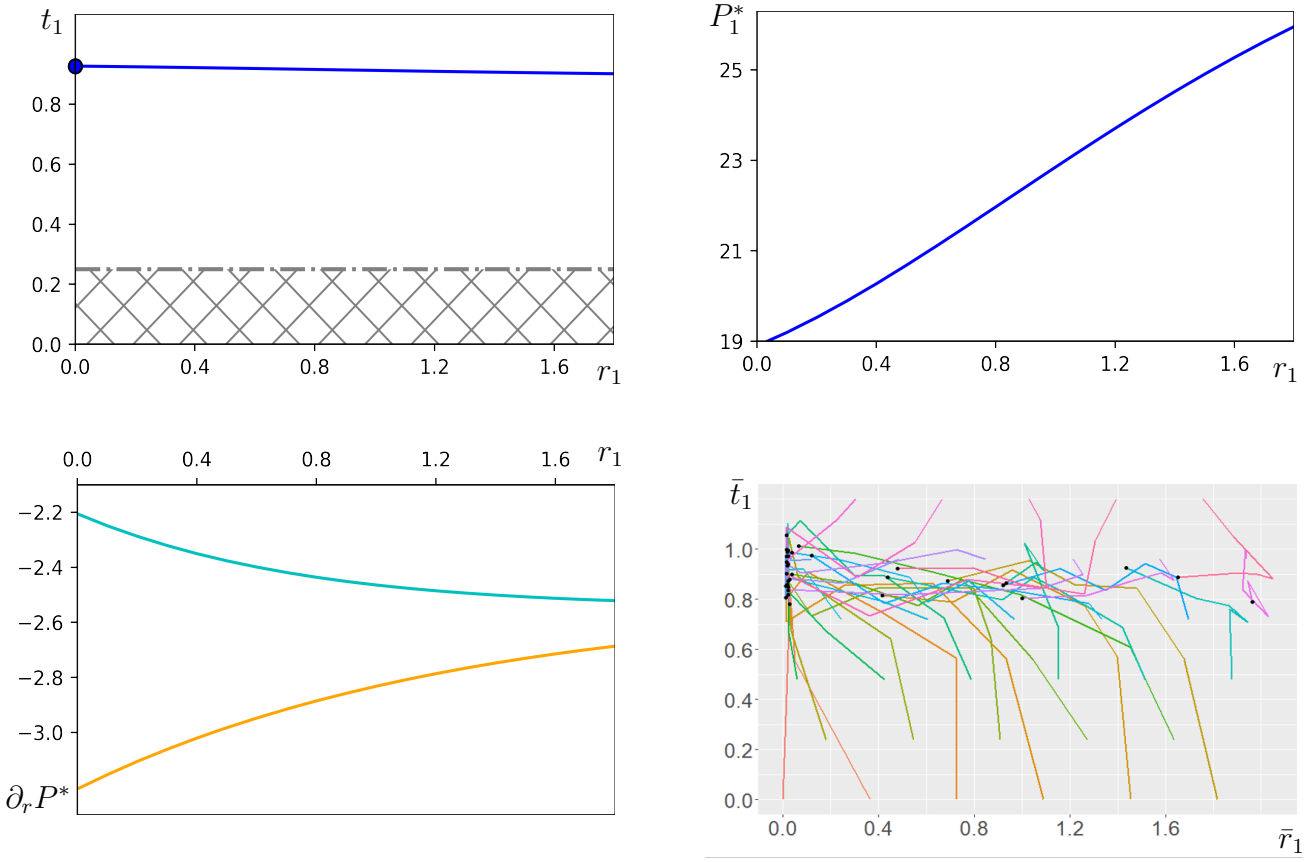

Figure 4.5: Parameter values $\lambda = 0.0015, a = 0, f = 5/6, k = 5, t_c = 0.25, q = 0.4, N = 100, n = 10, v = 1$ **(a)** [Top left] Unique cryptic ESS represented as a solid marker at $(0, 0.927)$ co-exists alongside a continuum of conspicuous ESSs shown in blue and defined implicitly through setting $a = 0$ into (4.1.2). Strategies (r_1^*, t_1^*) on the curve (4.1.2) are such that $t_1^* > c(r_1^*)$ so that the level t_1^* is predicted to decrease (slowly) with respect to r_1^*. **(b)** [Top right] Absolute resident fitness defined implicitly through (4.1.27) and plotted as a function of the conspicuousness. **(c)** [Bottom left] Invasion fitness gradient associated with less (cyan) and more conspicuous (orange) mutant types at equilibrium given through (4.1.15) and (4.1.16). **(d)** [Bottom right] Average population traits plotted as trajectories with averaging frequency $g = 2,000$. Black markers show strong converge to crypsis, which is mostly supported from Figures 4.5(b) and 4.5(c).

Before discussing Figures 4.5 and 4.6 individually, we should remark that these relate to the same example but where in Figure 4.6 different sets of trajectories are plotted for different levels of background mortality (Figures 4.5d and 4.6b are identical). In Figure 4.5(d) prey traits are mostly observed to converge to the equilibrium level shown in Figure 4.5(a), which is determined implicitly through setting $a = 0$ in (4.1.2). The level of defence in t_1^* in (4.1.2) is predicted to decrease with increasing levels of conspicuousness, although this effect is not captured in Figure 4.5(d) due to stochastic effects (for reasons discussed in due course these tend to be stronger when parameter a is small). The trajectories in 4.5(d) exhibit a strong pull

toward crypsis and this is more pronounced for lower values of the conspicuousness, where the cyan curve is highest. Presently, we confirm existing intuition (drawn from our discussions of Figures 4.2 and 4.3), namely that absolute resident fitness has limited impact on trait evolution when the local relatedness parameter is small/vanishing. Indeed, even in absence of strong directional selection, resident conspicuousness evolves against the resident fitness and toward lower values of $\bar{r}_1$.

These conclusions are valid for the remaining three plots in Figure 4.6, from which two additional conclusions can be drawn: As the background mortality increases the level of defence t_1^* decreases with respect to r_1^* in (4.1.2) and its relationship to r_1^* switches from decreasing (Figures 4.6a and 4.6b) to increasing (Figures 4.6c and 4.6d). The simulated plots in Figure 4.6 (this includes Figure 4.5d) exhibit considerably more randomness than their counterparts in Figure 4.7, which is likely attributed to the larger value of the local relatedness parameter in the latter (and its impact on initial invasion and fixation/drift). For this reason we elaborate on (i) and (ii) in the context of Figure 4.7 below and compare these to the analytical predictions in section 4.2.

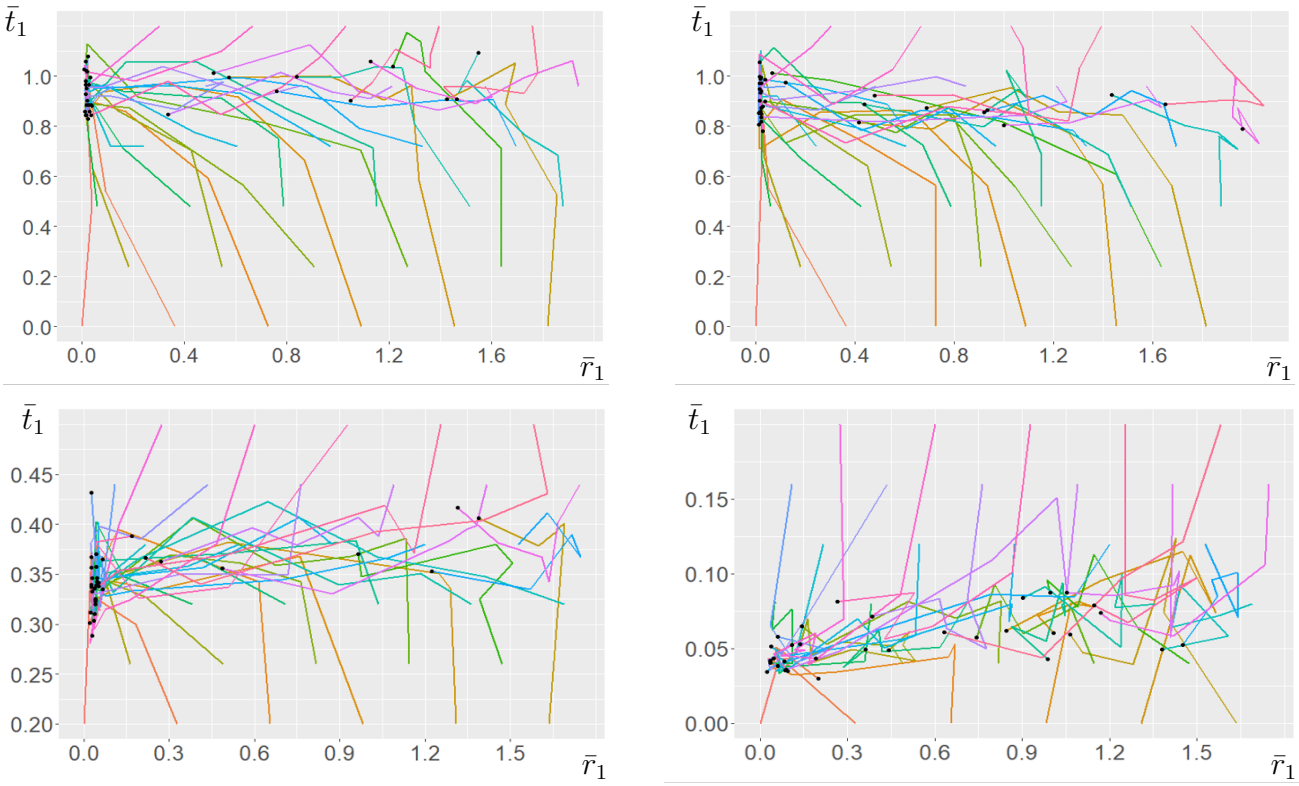

Figure 4.6: Parameters $a = 0, f = 5/6, k = 5, t_c = 0.25, q = 0.4, N = 100, n = 10, v = 1$ with plots in increasing order of the parameter λ with **(a)** [Top left] $\lambda = 0.0001$; **(b)** [Top right] $\lambda = 0.0015$; **(c)** [Bottom left] $\lambda = 0.2$ and **(d)** [Bottom right] $\lambda = 2$. Together the plots (mostly) confirm that increasing λ causes a decrease in the associated level of toxicity (for fixed conspicuousness) and that the relationship between t_1^* and r_1^* switches from decreasing (in 4.6a and 4.6b) to increasing (in 4.6c and 4.6d) as the predicted level of defence t_1^* drops below $c(r_1^*)$ as λ increases. The accumulation of black markers suggests strong selection for crypsis, likely driven by directional preference of mutant fitness in that direction. We also remark that trajectories in (a) convey a mostly flat equilibrium at $t \approx 1/f - 1/k$ and that in (d) trajectories are traced out within the non-aversive - and unstable in the sense of (4.1.15) - region $\bar{t}_1 \leq t_c = 0.25$.

Incorporating $a > 0$

We have established - this is done analytically in section 4.2 - that the level of defence at equilibrium decreases with increasing values of the parameter λ, such that for small values of λ, the associated level of defence is high enough that prey are highly aversive for predators. In such cases the plots in Figures 4.6/7 (a),(b) suggest that prey can increase aversiveness further by increasing conspicuousness while simultaneously (slightly) decreasing investment in defence. In contrast, if λ is high, the overall level of defence is low and prey are not very aversive so that larger conspicuousness selects for slightly more investment in defence (see Figures 6/7 c, d). The latter is likely because the gain in terms of (signalled) aversiveness outweighs the costs, which in turn can be traced back to with the choices of functional forms in (4.1.1).

The predicted slope of the equilibrium curve is provided by the *Implicit Function Theorem* in $\mathbb{R}^2$, which for the functions (**??**) used in the simulation takes the form shown in (4.1.5). As discussed in section 4.2 the single term in (4.1.2) can accommodate changes in monotonicity is $(\lambda/DKQ) \times F'/F$ and describes the impact on fecundity (of increased defence) scaled as a proportion of background to predator-induced deaths. This quantity can be seen as an honest measure for the capacity of investment in aposematic defences to increase prey fitness (through favourable trade-off involving life-span and reproduction). When λ is low (and prey are aversive) it is optimal for prey to increase their reproductive success by reducing their toxicity in favour of slightly higher mortality (seen through increased conspicuousness). The functional forms are such that when λ is high (and prey are non-aversive) the optimal trade-off regime changes so that it is best for prey to reduce their reproductive success (by increasing toxicity) in favour of reduced predation (seen through an increased conspicuousness). The reader is strongly encouraged to compare the findings of Figures 4.6/7 with the analysis in section 4.2.

Third, we remark that when the invasion fitness gradient along r is flat enough in either direction (i.e. more and less conspicuous mutants are worse-off alike) and provided the local relatedness parameter is strong enough, absolute resident fitness can have a notable influence on the evolution of prey traits. This is likely attributed to an amplification of the group effect that larger values of the parameter a has (see earlier explanation about frequency-dependence) and could explain why strategies with considerable signalling component are selected for when λ is sufficiently small (see Figures 4.7a and 4.7b). The latter is rather clearly showcased in Figure 4.7, where trajectories evolve against the invasion fitness gradient and toward increasing levels of the absolute resident fitness.

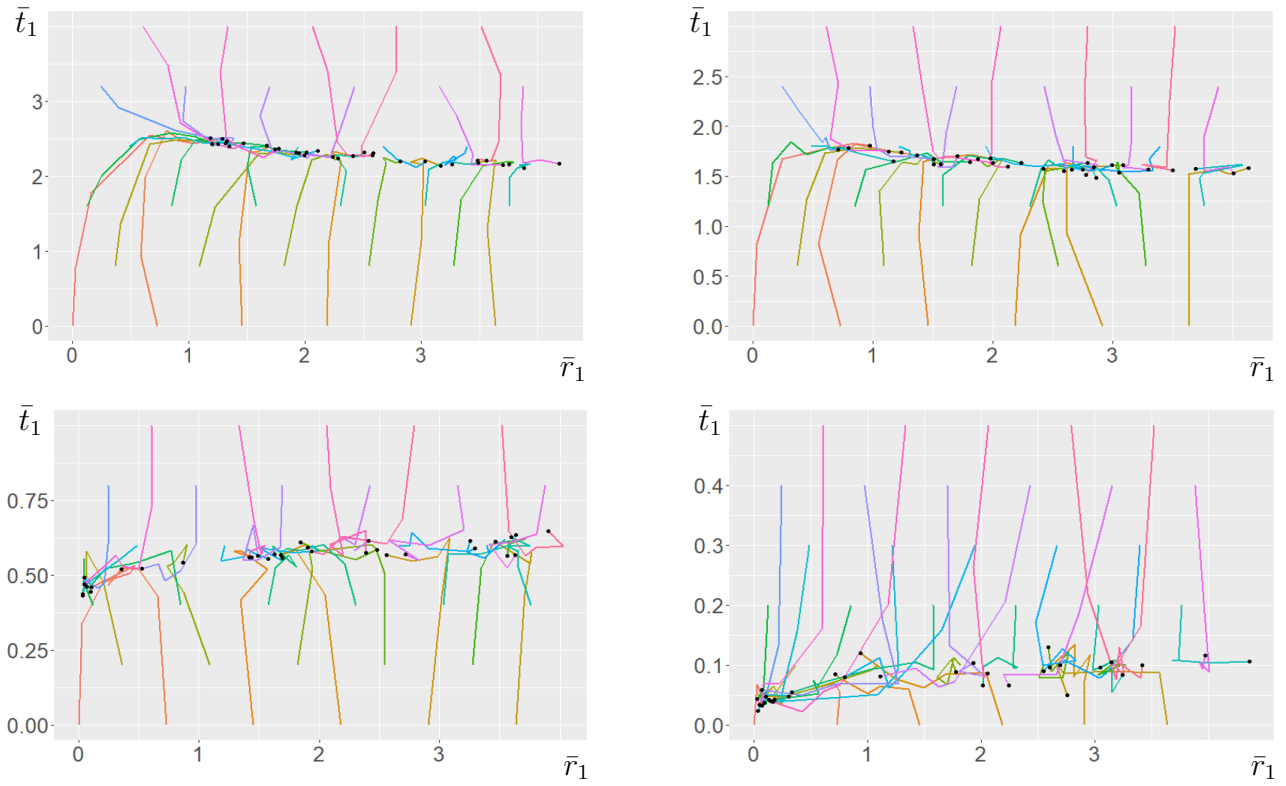

Figure 4.7: Parameter values $a = 0.5, f_0 = q_0 = k_0 = 1, f = 5/6, k = 5, t_c = 0.25, q = 0.4, N = 100, n = 10, v = 1$. Plots are positioned in increasing order of λ such that **(a)** [Top left] $\lambda = 0.0001$; **(b)** [Top right] $\lambda = 0.0015$; **(c)** [Bottom left] $\lambda = 0.2$ and **(d)** [Bottom right] $\lambda = 2$. The plots validate more clearly than in Figure 4.6 that increasing λ is associated with a decrease in the associated level of toxicity and that the relationship between toxicity and conspicuousness switches from negative (in 4.7a and 4.7b) to positive (in 4.7c and 4.7d). In Figures 4.7(a) and 4.7(b) there is strong selection for solutions with strong signalling component, likely on account of the absolute resident fitness being highest in that direction. In Figures 4.7(c) and 4.7(d) it is clear that the resident fitness is not sufficient to counterbalance the impact of a strong invasion fitness gradient (from the left and along r). Large spaces between black markers (such as at $\bar{r}_1 \approx 1$ in Figure 4.7c) are likely due to a balancing effect of these opposite pulls.

4.5 Discussion

The results presented in the previous section have demonstrated both the strengths and limitations of applying (infinite population) ESS analysis within the broader mathematical development of Broom et al. (2006) to study the evolution of prey traits in finite populations. To that end, had smaller populations been considered the outcomes of the simulations would have generally been driven by randomness. We should remark that spatiotemporal variations in the various environmental factors (including territory quality) and in the predator's community structure are not explicitly accounted for in our model, even though we acknowledge their importance in the selection for/against aposematism in real populations. For instance, as discussed in Mappes et al. (2005) the genetic predisposition and cultural transmission of foraging strategies within families could lead to strongly localised selection for/against aposematism. In this closing section we call attention to these points and argue that the simulation model as described in section 4.3 can be generalised to prey populations consisting of more than one species including Batesian mimicry complexes among others.

In an infinite population ESS analysis is all that matters, while in a very small population, stochastics

dominates. For intermediate numbers, stochastic mechanisms will eventually prevail in theory, but this may take a really long time, so effectively the ESS analysis is indeed all that is needed. In the t-direction, in any mixed population where all prey have similar conspicuousness, the optimal toxicity level is approximately the same, independent of the precise composition of the population, as long as the average conspicuousness does not change too much, or indeed often even if it does). In the r-direction, we often have a series of populations that are stable, but where the neighbouring mutants are not so much worse, so there is the prospect of invasion due to chance. It is when we get to a substantial mutant sub-population that the mutants "resident fitness" (i.e., the mutants' fitness after fixation) comes into play. Indeed full invasion is more likely to happen from the higher fitness side, so there will tend to be movement in that direction. The latter is manifest in Figure 7, where when the background mortality is sufficiently low aposematic strategies with considerable signalling component are selected (likely due to the higher associated fitness).

In this manuscript we have considered the functional forms of Broom et al. (2008) and compared regimes with and without background mortality. The conclusions drawn in previous works on aposematism have been constrained by the assumption that predation is the only source of prey death and the impact of varying regimes of background mortality has prior to now not been explored. In addition to accounting for sources of prey death outside of predation, we have explored the effect of local clustering through parameter a and utilised ESS (and fitness) analysis to draw conclusions about the evolution of prey traits in intermediate populations that are subject to stochasticity. While simulation models have been used in G. D. Ruxton and Beauchamp (2008) and in Zakharova et al. (2019) and elsewhere over the recent decades, these have never before been put to use to study aposematism. We have made contributions to the game-theoretical model of Broom et al. (2006) by broadening the scope of ESS analysis, by implementing it into a novel simulation model and by gauging the capacity of ESS (and fitness) analysis to predict the evolution of aposematism in finite prey populations.

In Broom et al. (2008) and previously in Broom et al. (2006) the use of simple functional forms and the suppression of background mortality had allowed to express the toxicity explicitly in terms of the conspicuousness and to conjecture that more conspicuous appearances are associated with prey that are better-defended. This conjecture was disproved in Scaramangas and Broom (2022) and also presently, where a decreasing relationship between conspicuousness and defence was observed (see simulations in Figure 4.7) in regimes where prey death outside of predation is rare. In Scaramangas and Broom (2022) as well as in chapter 3 the justification of a decreasing relationship involved the implementation of a more elaborate (plausible nonetheless) association between the predator's propensity to attack based on its perception of prey aversiveness (through a modification of the form for Q). Interestingly, such a modification had also allowed us to disprove another conjecture of Broom et al. (2006) and to demonstrate that a certain level of signal strength may be associated with more than one ESS level of the defence.

Although the observation of a decreasing signalling-defence continuum is in this manuscript linked with functional forms that are different to those of Scaramangas and Broom (2022), perhaps the underlying mechanism is common. In the first simulations of Figure 4.7 (a and b) we observe that when the overall ESS level of defence is sufficiently high (background mortality is low) prey can afford to broadcast weaker defences through stronger signals because predator propensity to attack is already low and saturated. This is also complemented by the fact that a further investment in toxicity is costly to the fecundity and this is a cost worth bearing if it is manifest through reductions in predation, which in this case is not. In contrast, when background mortality is high and the associated toxicity is low (see simulations in Figures 7c and d) brighter appearances signal stronger defences because the reduction in fecundity is compensated with a

reduction in predation and an increase in average prey life-span.

Indeed, of considerable importance to the theory of aposematic signalling is whether aposematic signals are *honest* (i.e. whether brighter prey are better defended) and the reader is encouraged to consult the review article by Summers et al. (2015) for a thorough account of this topic. While there is more empirical evidence reporting a positive relationship between conspicuousness and defence (Summers and Clough, 2001b, Santos and Cannatella, 2011 and Maan and Cummings, 2012b are among several cited in Summers et al., 2015) there are noteworthy studies (including those of I. J. Wang, 2011 and Darst et al., 2006) suggesting that conspicuous signals could be dishonest. As argued in Scaramangas and Broom (2022) the model of Broom et al. (2006) is the only detailed exposition that can account for the full breadth of phenomena and this is observed presently.

The theory presented here makes clear predictions that would allow empirical testing. Perhaps our more interesting predictions stem from the comparison between the analytic theory and the simulations. It seems clear that when prey populations are large then the predictions of both modelling approaches converge, but for smaller populations the stochastic fluctuations captured in the simulation model should have a strong bearing. It would be valuable to explore experimentally with living prey how small a population has to be for these stochastic effects to have a strong bearing on evolutionary trajectories, how strong these effects are, and how exactly they alter the course of evolution. It seems more easy to imagine how such empirical explorations could be achieved in the laboratory than in natural populations. But even here there will be a challenge in finding a suitable prey type that can readily be kept in large numbers and shows the combination of appearance and toxicity characteristics of interest to us and that has a short enough generation time that meaningful evolutionary trajectories can be followed. A candidate here might be one of the stored-product beetles that are increasingly becoming model species for studies in evolution and population dynamics (of much relevance is the review article by Pointer et al., 2021). The most commonly-used species in such studies (Tribolium castaneum) is chemically defended and shows variation in coloration from red, through browns to black (see McLean, 2011).

We imagine that such experiments would involve not natural predators but artificial predation imposed by the experimenters – with different types of predation represented by removal of prey individuals from the population as defined by different sets of rules (mimicking the assumptions about predator behaviour in our theory). As well as exploring the consequences of prey population size (and indeed the size of the artificial predator population – as represented by the intensity of predator-mimicking mortality) on evolutionary trajectories – it would be straightforward to also explore our predictions about the effect of additional external non-predatory mortality in these experiments.

We also think that experiments with real predators would also be valuable in the context of testing our predictions. Well-developed systems for investigating why predators learn about aposematic prey and how this affects subsequent prey choice decisions already exist. These can use completely prey-naïve newborn domestic chicks (as in Rowland et al., 2013) or wild-caught insectivorous birds temporarily exposed to artificial prey in a laboratory setting (such as in Hämäläinen et al., 2020). Our model assumptions and predictions related to how predators respond in successive encounters with different types of prey items – particularly the assumptions about the spatial distribution of mutant types encapsulated in our parameter (a) could very naturally be explored empirically with such a system.

Furthermore, we see value in co-evolutionary experiments that allow us to explore whether the assumptions we make for predator behaviour in our models are likely a reasonable representation of those that evolve in real predators. For this, we might return to the evolutionary experiments with a simple laboratory

prey organism like stored flour beetles discussed above, but rather than subjecting them to an unchanging predation regime, we allow the predatory regime to co-evolve with the prey. We have in mind here a population of artificial predators – each of which follows a set of rules about how it treats prey of different types, and thus imposes mortality on the prey population. However, variation in these rules will not only lead to variation in the form of mortality imposed on the prey but also on the fitness of the artificial predators – where a fitness score is awarded according to how well the predator exploits lower-defended prey and avoids higher-defended prey. If at each generation of the real prey the artificial predator population is changed such that more successful rule-systems become more prevalent in the artificial predator population, then we can effectively mimic predator-prey co-evolution – and most pertinently we can explore whether the predator population coalesces to rules that have commonality with those assumed in our theory. There is a collection of interesting studies examining the co-evolution of aposematic prey in a prey-predator complex including Teichmann et al. (2014b) and Teichmann et al. (2015), whose results may be of particular insight to the experimenter.

We have seen how stochastic mechanisms can impact the evolution of aposematism in finite prey populations to varying degrees. While our conclusions about the evolution of prey traits in regimes (i), (ii), (iii) and (iv) cannot be considered generic, the choice of parameter values is important and related to the size of the mutation step. Throughout the simulations we have assumed a (maximum) mutation step of fixed size (the same value regardless of the resident strategy from which mutation is occurring) so that depending on the regime considered, different sets of parameter values can lead to different predicted outcomes (seen through the level of t_1^*). In every regime the exact choices of values were gauged manually so as to mitigate the impact of stochasticity. While this method proved effective, it is by no means optimal. As such, future efforts could consider a large number of identical systems (identical in the sense of sharing the same parameter values) and showcase the time evolution of prey traits as an average over these. This type of averaging is known as an *ensemble average* and is especially effective for mitigating the impact of stochasticity (likely more so than the averaging process used currently in individual runs of the simulation). [1]

In closing, we would like to highlight the success of the simulation in showcasing the evolution of aposematism in prey populations that are finite. We would also like to argue that it is possible to extend the game-theoretic treatment of Broom et al. (2006) to account for Batesian mimicry systems, which are arguably among the most important (and most studied) mimicry complexes encountered in nature. Work of this type could utilise the territorially-divided habitat structure referred to in Scaramangas and Broom (2022) and introduce on this a proportion of (beta-distributed) undefended mimics. Achieving stability of a model and a mimicking species in a certain habitat on the (longer) evolutionary time-scales requires that the individual sub-populations are stable on the (shorter) ecological time-scales and such a condition need be considered jointly with the ESS conditions detailed here. Research in this direction is promising and currently underway.

4.6 Code

In this section of the appendix we include a sample of the code that was used in R to run the simulation in Figure 4.4(d). The plots for the remaining simulations were generated in a similar fashion.

[1] Ensemble averages were originally used in statistical physics to describe the averages of quantities that depended on the *microstate* (e.g. the individual configuration of the components) of the system. They are now used extensively in other fields outside of statistical physics, such as in machine learning, weather forecasting and signal processing.

```
### Figure 4(d) ###

lamda <- 0
a <- 0.5

nprey <- 100
npred <- 10

frac_die <- 0.5 # The fraction of the population dying at the end of each iteration
ndie <- trunc(nprey*frac_die)

f0 <- 1
d0 <- 1
k0 <- 1
q0 <- 1
f <- 2.8
k <- 5
tc  <- 0.25
q <- 0.4
v <- 1

iterations <- 10000

rmin <- 0
rmax <- 5
rdiff <- rmax-rmin

tmin <- 0
tmax <- 1.5
tdiff <- tmax-tmin

listt <- c(rep(tmin,6), rep(tmin + tdiff/5,5), rep(tmin + 2*tdiff/5,6), rep(tmin + 3*tdiff/5,5),
rep(tmin + 4*tdiff/5,6), rep(tmax,5))

listr <- c(seq(rmin, rmin + 10*rdiff/11, rmin + 2*rdiff/11),
seq(rmin + rdiff/11, rmin + 9*rdiff/11, rmin + 2*rdiff/11),
seq(rmin + (1/3)*rdiff/11, rmin + (10+1/3)*rdiff/11, rmin + 2*rdiff/11),
seq(rmin + (1+1/3)*rdiff/11, rmin + (9+1/3)*rdiff/11, rmin + 2*rdiff/11),
seq(rmin + (2/3)*rdiff/11, rmin + (10+2/3)*rdiff/11, rmin + 2*rdiff/11),
seq(rmin + (1+2/3)*rdiff/11, rmin + (9+2/3)*rdiff/11, rmin + 2*rdiff/11))

progt <- c()
progr <- c()
```

```
for(d in 1:length(listt)){

  meant <- c(rep(0, iterations))
  meanr <- c(rep(0, iterations))
  iterrs <- c(rep(seq(1,iterations,1),2))

  progt <- append(progt, listt[d])
  progr <- append(progr, listr[d])

  # Initialise individual prey properties

  P <- c(rep(0,nprey))
  D <- c(rep(0,nprey))
  K <- c(rep(0,nprey))
  Q <- c(rep(0,nprey))
  I <- c(rep(0,nprey))

  t <- rep(listt[d], nprey)
  r <- rep(listr[d], nprey)

  tstart <- t
  rstart <- r

  rsum <- 0
  tsum <- 0
  iters <- 0

  print(d)

  for (iter in 1:iterations){

    meant[iter] <- mean(t)
    meanr[iter] <- mean(r)

    if(iter%%(iterations/5)<0.5){

      avt <- sum(meant[(iter-(iterations/5-1)):iter])/(iterations/5)
      avr <- sum(meanr[(iter-(iterations/5-1)):iter])/(iterations/5)

      progt <- append(progt, avt)
      progr <- append(progr, avr) }
```

```
# Determine prey fitness

F <- f0*exp(-(f*t))
D <- 1/(1+exp(-r))
K <- k0/(1+(k*t))

for (x in 1:nprey){
  sum <- 0
  for (y in 1:nprey){
    if ((abs(x-y)) > 0.01){
      L <- D[y]
      H <- t[y] - tc
      S <- 1-(v*(abs(r[x]-r[y])))
      if (S<0){S<-0}
      sum <- sum +(L*H*S)}}

  L_self <- D[x]
  H_self <- t[x] - tc
  S_self <- 1

  I[x] = (((nprey*(1-a)/(nprey-1))*sum)+((a*nprey-1)*L_self*H_self*S_self))/npred}

Q <- q0*exp(-q*I)
for (z in 1:length(Q)){
  if (Q[z]>1){Q[z]<-1}}

P <- F/(lamda+(D*K*Q))


# Update population w.r.t. fitness

wts_reproduce   <- P/sum(P)

wts_die <- 1-wts_reproduce
sumwts <- sum(wts_die)
wts_die <- wts_die/sumwts

parents <- sample(nprey, size = ndie, replace = TRUE, prob = wts_reproduce)
replaced <- sample(nprey, size = ndie, replace = FALSE, prob = wts_die)

copyr <- r
copyt <- t
```

```
    for (x in 1:ndie){
      p1 <- parents[x]
      q1 <- replaced[x]
      mutr <- rbinom(1,1,0.05)
      if(mutr>0.5){
      r[q1] <- copyr[p1]+((-0.05)+(0.1*runif(1)))
        if(r[q1]<0){r[q1] <- 0}}
      else{r[q1] <- copyr[p1]}
      mutt <- rbinom(1,1,0.05)
      if(mutt>0.5){
        t[q1] <- copyt[p1]+((-0.05)+(0.1*runif(1)))
        if(t[q1]<0){t[q1] <- 0}}
      else{t[q1] <- copyt[p1]}} # End of change
      }
# End of iteration
}


# Produce plot for Figure 4(d)

length(progr)
length(progt)

lines <- rep(1:33, each=6)
lines <- as.factor(lines)

idealfig <- data.frame(progr, progt, lines)

markersr <- progr[seq(0, length(idealfig$progr), 6)]
markerst <- progt[seq(0, length(idealfig$progt), 6)]

markers <- data.frame(markersr, markerst)

plot4d <- ggplot(idealfig, aes(x=progr, y=progt)) +
  geom_path(aes(group=lines, col=lines), size=1.05) +
  theme(legend.position="none", axis.text.x = element_text(size=20),
  axis.text.y = element_text(size=20))
  + xlab("") + ylab("") +
  geom_point(data=markers, mapping=aes(x=markersr, y=markerst))+
  scale_x_continuous(breaks=seq(0,4,1))+
  scale_y_continuous(breaks=seq(0,1.2,0.3))
```

```
plot4d

# End of code
```

Chapter 5

Co-existence & mimicry: a first approach

The purpose of this and the following chapter are to extend the model of Broom et al. (2006) to account for instances in which the prey population is made up of either two types belonging to the same species (but playing different strategies) or to instances in which the prey population is made up of two distinct species altogether. Although from an evolutionary point of view these two cases are quite different, we demonstrate presently and in chapter 6 that from the mathematical modelling perspective these can be considered jointly. Indeed, we develop the conditions that are required for two types to co-exist in a sense that is stable both from the ecological and from the evolutionary point of view (we call this *eco-evolutionary stability*). Following the approach of Broom and Rychtár (2013) we treat the time-scale in which the population dynamics evolve to be notably shorter than the time-scale in which evolutionary processes occur, such that when considering ecological stability we imagine that the evolutionary processes are fixed and vice versa, when accounting for evolutionary stability we imagine that the population dynamics are in equilibrium. We should also remark that while ecological stability is considered alongside the ESS analysis, the treatment remains static and independent of specific assumptions regarding the dynamics.

So as to focus the attention to those aspects of the mathematical modelling that are novel, we make a number of simplifying assumptions during the numerical analysis portion of the chapter, which we consider plausible. For instance, we keep with the assumption that predation is the only source of prey mortality [1] or that predators have perfect recollection of encounters (so that regardless of whether a detected individual is attacked and/or captured the encounter is remembered). The structuring of this chapter is similar to the structuring of the sixth chapter and is as follows: first, we explain how the setup described in chapter 2 is extended to account for a two-type prey population; second, following this we discuss what it means for the types to co-exist in a sense that is eco-evolutionarily stable (and express the conditions in terms of general functional forms); third, we identify by means of elimination the admissible forms for the function Q describing the probability of attack that are best-suited for studying mimicry under the specific assumptions of this chapter. In closing, we provide a working example of a system that is eco-evolutionarily stable, initially in the form of a point solution (i.e. exact values for the type-1 and type-2 strategies and proportions) and hence extend the method detailed therein to demonstrate how one can recover a continuum of solutions (with the type-1 and type-2 strategies and proportions drawn from intervals of values). The mathematical

[1] Indeed, it was demonstrated in chapter 4 that regimes with non-zero rates of background mortality can be inherently more challenging to model, even for the case in which the prey population consists of a single species. Accounting for sources of death outside of predation within a two-type prey population and the possibility of testing these predictions against a genetic algorithm model (see section 4.3) provide a potentially fruitful future objective.

modelling of mimicry is strikingly limited in the literature and for this reason we have decided to provide more detail in the calculations than in the previous chapters and to include specific workings for the numerical analysis.

5.1 Two-type co-existence

While the motivation behind this chapter is mimicry, the presentation in this first section does not impose a specific relationship between the strategies of one or the other type and may in principle be considered as a basis from which to explore a wider variety of co-existence problems involving three protagonists: a predator and two prey types. That said, due to time resources being limited we consider co-existence only on a preliminary level and do not explore explicit numerical examples beyond the strictly mimetic cases. Indeed, while the discussion is kept as general as possible wherever possible, the larger portion of this chapter is devoted exclusively to co-existence in the form of mimicry. This is achieved by interpreting the first type as the *model* (consisting of aversive prey emitting signals of given conspicuousness) and the second type as the *mimic* who resembles the model and is completely undefended (although we do include cases where this can still be defended - see Definition 5.1.1 and Theorem 5.1.2). In the worked examples that are provided later in the chapter (see sections 5.3 and 5.4) we discuss mimicry in a rather idealised context of *perfect resemblance* in which the mimic is completely indistinguishable to the model from the point of view of the predator.

Payoff and perceived aversiveness

We imagine an extended habitat structure that is territorially-divided among the predators such that each site consists of N prey and is visited by n predators who visit that site only. In the $\varepsilon \to 0$ limit of the resident-mutant description provided in chapter 2 it was assumed that almost all sites consist of prey playing the resident strategy (r_1, t_1) except for a small, effectively negligible number of sites containing clusters (of size a) of copies of a certain focal individual. In keeping with the $\varepsilon \to 0$ limit description, we presently extend the latter in a straightforward way by assuming that in almost all sites can be found a proportion $(1-\gamma)$ of prey of a certain type (type-1) playing strategy (r_1, t_1), with remaining prey (type-2) playing strategy (r_2, t_2) in proportion $\gamma \in [0, 1]$.

We impose mutation by imagining that while in almost all sites there is a $(1-\gamma)/\gamma$ split between type-1s and type-2s that there is a small number of sites containing clusters of focal relatives in which this proportion is perturbed. [2] Under this description we account for local mutation in the traits of either the first or the second type by maintaining that the mutant trait (r, t) can assume values that are local to either the type-1 or the type-2 strategy (see Definition 5.1.1 for a more precise description of *local* based on where on the strategy space each type draws its strategy from).

In both this and in chapter 6 we conceive that mutation is facilitated by means of local clustering, so that while mutants make up a negligible proportion of the overall prey population their presence on the local level can have considerable impact on their aversiveness as perceived by the group of predators visiting their site. In this chapter in particular, the local relatedness parameter a is understood as the proportion of relatives of the (focal) individual who play strategy (r, t) *over the total number of prey residing in that site*.

[2] We should remark that the limiting cases $\gamma = 0$ and $\gamma = 1$ describe situations where either only the first type or only the second type is present and therefore does not strictly describe co-existence. This remark is more relevant when it comes to considering specific outcomes; for now it is safe to assume that γ assumes value on the closed interval $[0, 1]$.

This interpretation of the relatedness is quite different to the one provided in chapter 6, in which the focal individual is of either one or the other type and in which that parameter measures the local concentration of relatives *as a proportion over the number of individuals of the focal type* - see section 6.1. In addition, while in this chapter γ is understood as a background proportion of type-2s (i.e. almost every site has this proportion) in chapter 6 we model this as a (beta-distributed) continuous random variable so that one expects to see different type-2 individuals in varying proportions in the habitat. We motivate the possibility of an alternative approach to modelling co-existence in the end of this chapter.

The interpretation of relatedness as used in this chapter suggests that in a site that does contain relatives there are aN relatives playing (r,t), while from the remaining $(1-a)N$ prey in that site there are $(1-a)(1-\gamma)N$ type-1s playing (r_1,t_1) and $(1-a)\gamma N$ type-2s playing (r_2,t_2) - strictly speaking the focal individual itself should be discounted but discussions of this type have already been made. From the perceived aversiveness as defined on the level of the individual in (2.2.2) as

$$I_i = \frac{1}{n}\sum_{j=1,j\neq i}^{N} L(r_j)H(t_j)S(|r_i - r_j|) \tag{5.1.1}$$

we recover the expression for the aversiveness of the focal individual as perceived by the group of predators visiting the local area as

$$\begin{aligned} I' &= a\frac{N}{n}L(r)H(t) + (1-a)(1-\gamma)\frac{N}{n}L(r_1)H(t_1)S(|r-r_1|) + (1-a)\gamma\frac{N}{n}L(r_2)H(t_2)S(|r-r_2|) \\ &= a\mathcal{I} + (1-a)(1-\gamma)\mathcal{I}_1 S(|r-r_1|) + (1-a)\gamma\mathcal{I}_2 S(|r-r_2|). \end{aligned} \tag{5.1.2}$$

This level of perceived aversiveness is to be contrasted with the perceived aversiveness of a type-1 individual, which is recovered through (5.1.2) by setting $(r,t) = (r_1,t_1)$ and reads

$$\begin{aligned} I_1' &= [a + (1-a)(1-\gamma)]\frac{N}{n}L(r_1)H(t_1) + (1-a)\gamma\frac{N}{n}L(r_2)H(t_2)S(|r_1-r_2|) \\ &= [a + (1-a)(1-\gamma)]\mathcal{I}_1 + (1-a)\gamma\mathcal{I}_2 S(|r_1-r_2|). \end{aligned} \tag{5.1.3}$$

Likewise, the perceived aversiveness of a type-2 individual is recovered through (5.1.2) by setting $(r,t) = (r_2,t_2)$ and reads

$$\begin{aligned} I_2' &= (1-a)(1-\gamma)\frac{N}{n}L(r_1)H(t_1)S(|r_1-r_2|) + [a + (1-a)\gamma]\frac{N}{n}L(r_2)H(t_2) \\ &= (1-a)(1-\gamma)\mathcal{I}_1 S(|r_1-r_2|) + [a + (1-a)\gamma]\,\mathcal{I}_2. \end{aligned} \tag{5.1.4}$$

We make use of the quantities $\mathcal{I}$, $\mathcal{I}_1$ and $\mathcal{I}_2$ to describe the associated level of aversive information if the habitat were made up entirely of a focal individuals, type-1s or type-2s respectively such that

$$\mathcal{I}_{1,2} := \frac{N}{n}L(r_{1,2})H(t_{1,2}). \tag{5.1.5}$$

Thus far, we have referred to the groups making up the prey population as *types* and have (purposely) left unspecified whether these belong to the same or to different species. From the mathematical modelling

perspective there are a number of items to consider on this front. The description of Broom et al. (2006) is one that models aposematism in prey and is not intended to explicitly model predator learning. In keeping with this, we use a single set of functional forms for L, H, S and Q to describe the predator's perception of prey as we do in (5.1.2), (5.1.3) and (5.1.4), as opposed to two sets. That is to say, we do not use one functional form for say Q to evaluate the perceived aversiveness of a type-1 and a different functional form for evaluating the aversiveness of a type-2 individual. Incorporating a second type to the prey population in such a manner we keep the predator parameters with respect to the learning process fixed.

On the other hand, we account for inherent differences between the types of prey by making use of different sets of functional forms to describe prey-related processes. Investment in secondary defences is costly to the individuals that deploy them and the extent to which this influences their fecundity should in principle depend on the type of prey in question. For this reason we consider two functional forms F_1 and F_2 to describe the relationship to defence of the fecundity for type-1s and type-2s. We could also argue that the extent to which a given level of investment in defence is effective at reducing that prey's chance of escaping an attack is also an inherent property of the prey in question and use forms K_1 and K_2 to distinguish between the types; in the chapter that follows we indeed use two sets of functions for both F and K. In this chapter we distinguish between the two types of prey only through F. The reader is also encouraged to review the related discussion in chapter 3 where in (3.2.7) was introduced parameter b as a measure describing the sensitivity of prey to investment in secondary defences that included both F and K; interestingly, it was shown that the outcomes for ESS exhibit strong dependence on this parameter.

Since the two types exhibit different sensitivity to investment in defence (through F) we introduce the payoff to a type-1 individual as

$$P_1(r_1, t_1; r_2, t_2) = \frac{F_1(t_1)}{\lambda + D(r_1)K(t_1)Q(I_1')} \tag{5.1.6}$$

and the payoff to a type-1 mutant as

$$P_1^{\dagger}(r, t; r_1, t_1; r_2, t_2) := \frac{F_1(t)}{\lambda + D(r)K(t)Q(I')} \tag{5.1.7}$$

with (r, t) defined in the local vicinity of the type-1 strategy (r_1, t_1). Likewise, the payoff to the second type is given as

$$P_2(r_1, t_1; r_2, t_2) := \frac{F_2(t_2)}{\lambda + D(r_2)K(t_2)Q(I_2')} \tag{5.1.8}$$

and for a mutant playing a strategy local to the type-2 strategy the payoff is given as

$$P_2^{\dagger}(r, t; r_1, t_1; r_2, t_2) := \frac{F_2(t)}{\lambda + D(r)K(t)Q(I')}. \tag{5.1.9}$$

For completeness, we remark that the mutant fitnesses are continuous at the value of each type on account of (5.1.2). Indeed, if $(r, t) = (r_1, t_1)$ we recover that $I' = I_1'$ so that

$$P_1^{\dagger}(r = r_1, t = t_1; r_1, t_1; r_2, t_2) = \frac{F_1(t_1)}{\lambda + D(r_1)K(t_1)Q(I_1')} = P_1(r_1, t_1; r_2, t_2). \tag{5.1.10}$$

Likewise, it is true that if $(r, t) = (r_2, t_2)$ that $I' = I_2'$ so that the mutant fitness is continuous at the type-2

strategy

$$P_2^\dagger(r=r_2,t=t_2;r_1,t_1;r_2,t_2)=\frac{F_2(t_2)}{\lambda+D(r_2)K(t_2)Q(I_2')}=P_2(r_1,t_1;r_2,t_2). \tag{5.1.11}$$

In chapter 2 it was discussed that the functional forms for F, D, K, Q, L and H are (by construction) of class $\mathcal{C}^l$ with $l \geq 2$ and the same applies to S sufficiently near the origin. The mutant fitness functions $P_1^\dagger$ and $P_2^\dagger$ as defined in (5.1.7) and (5.1.9) are composed of such functions and of the bi-variate similarity function $\mathcal{S}(r_i, r_j) = S(|r_i - r_j|)$ - see definition in (2.2.5) - which depends only on the conspicuousness trait and is not differentiable with respect to r_i or r_j at $r_i = r_j$. From the above we draw the following conclusions about the fitness functions $P_1^\dagger$ and $P_2^\dagger$. From this it follows that the mutant fitness functions are almost everywhere $\mathcal{C}^l$ with $l \geq 2$ except at $r = r_1$ and $r = r_2$. The latter is much like in (2.3.5) of chapter 2, where it was shown that the mutant fitness is non-differentiable along r at $r = r_1$; it now follows that the mutant fitness functions $P_1^\dagger$ and $P_2^\dagger$ are non-differentiable (but are continuous) at $r = r_1$ and at $r = r_2$ respectively. Indeed, we have

$$\begin{aligned} C_{1,2}(r,t)\partial_r P_{1,2}^\dagger(r,t;r_1,t_1;r_2,t_2) = &-\frac{D'(r)}{D(r)}-a\mathcal{I}\frac{L'(r)}{L(r)}\frac{Q'(I')}{Q(I')} \\ &-(1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I')}{Q(I')}S'(|r-r_1|)\Big(-\mathbb{1}_{(-\infty,r_1)}+\mathbb{1}_{(r_1,\infty)}\Big) \\ &-(1-a)\gamma\mathcal{I}_2\frac{Q'(I')}{Q(I')}S'(|r-r_2|)\Big(-\mathbb{1}_{(-\infty,r_2)}+\mathbb{1}_{(r_2,+\infty)}\Big) \end{aligned} \tag{5.1.12}$$

with

$$C_{1,2}(r,t)=\frac{(\lambda+D(r)K(t)Q(I'))^2}{F_{1,2}(t)D(r)K(t)Q(I')}$$

as per usual for the scaling factor. In Theorem 5.1.2 of the following section we discuss how fitness functions with these properties can be used to determine when the prey population is (locally) evolutionarily stable. Besides providing evidence that the mutant payoffs $P_1^\dagger$ and $P_2^\dagger$ are not differentiable along r at $r = r_1$ and $r = r_2$ the RHS of (5.1.12) suggests that the relative ordering of the quantities r_1 and r_2 is determining; such that cases with $r_1 < r_2$, $r_1 > r_2$ and $r_1 = r_2$ need be considered separately (this is discussed in greater detail at the end of this section, following the proof of Theorem 5.1.2).

Fitness is in this manuscript understood as a measure describing the number of offspring produced per life cycle so that high-fitness individuals are those who produce many offspring and/or those with longer life-cycles. This interpretation has allowed us to draw conclusions about the future composition of a population, based on its present composition using an approach that is static. For instance, we contend that if the overall fitness to a type-1 is greater than the overall fitness to a type-2 that the population of the first type will eventually outgrow the population of the other. While this process is not explicitly modelled it is presumed to take place on a time-scale that is shorter than the time taken for mutation to occur and potentially disrupt the composition.

Ecological stability

There is perhaps little sense in determining whether or not two types can co-exist on the longer time-scale if from a fitness perspective one type has the potential to outgrow the other. In keeping with an approach that is exclusively static we impose a set of conditions (separate to ESS) on the fitnesses of each type to

define what we describe as *ecological stability* in this and in the following chapter. In particular, we say that two types playing strategies (r_1, t_1) and (r_2, t_2) are in *ecological equilibrium* if when played in proportions $(1-\gamma^*)$ and γ^* - providing there exists such a proportion - the associated fitnesses of each type are equal

$$P_1(r_1, t_1; r_2, t_2) = P_2(r_1, t_1; r_2, t_2), \tag{5.1.13}$$

which from (5.1.6) and (5.1.8) amounts to

$$\frac{F_1(t_1)}{\lambda + D(r_1)K(t_1)Q(I_1')} = \frac{F_2(t_2)}{\lambda + D(r_2)K(t_2)Q(I_2')}. \tag{5.1.14}$$

From the definitions of the quantities I_1' and I_2' provided in (5.1.3) and (5.1.4) it is clear that the fitness of either type is sensitive to their relative abundance. Indeed, it is possible for two types to satisfy (5.1.14) for some proportion γ^* but for the composition to be violated when the proportion is perturbed marginally away from the equilibrium level. To eliminate this prospect, we impose that the equilibrium proportion is stable under such perturbations by means of the following inequality

$$\partial_\gamma \Big(P_2(r_1, t_1; r_2, t_2) - P_1(r_1, t_1; r_2, t_2) \Big) \Big|_{\gamma=\gamma^*} < 0. \tag{5.1.15}$$

The reader is directed to the classic Hawk-Dove game analysis in chapter 2 of Maynard Smith (1982) for more insight into what we presently describe as ecological equilibrium. The discussion there, as well as in chapter 4 of Broom and Rychtár (2013) - see Figure 4.2 therein - includes an analysis of the behaviour of a population as it deviates from a state of polymorphic equilibrium. It is shown that such deviations (due to stochastic effects or due to the finiteness of the population) lead to a reduction in (average) fitness so that polymorphism is restored and eventually the population evolves toward the ESS strategy.

Inequality (5.1.15) guarantees that if the types play (r_1, t_1) and (r_2, t_2) in a proportion that is marginally smaller than the equilibrium value given through (5.1.14) that this will be compensated by a marginally larger type-2 fitness that acts to restore the population back to the equilibrium proportion. Likewise, if the types co-exist in proportion that is marginally larger than the equilibrium level this is compensated by a marginally larger type-1 fitness that tends to restore the population to the equilibrium proportion in (5.1.14). From context it should be clear that the sign of the term in brackets ("type-2 fitness minus type-1 fitness") in (5.1.15) is as such on account of the definition of γ as the proportion of type-2 individuals. Indeed, if the direction of the inequality were reversed, a negative deviation away from the equilibrium proportion would lead to the extinction of the type-2 population ($\gamma \to 0$), while a positive deviation away from the equilibrium proportion would lead to the extinction of the type-1 population ($\gamma \to 1$).

Substitution of (5.1.6) and (5.1.8) into (5.1.15) leads to

$$-\frac{F_2(t_2)D(r_2)K(t_2)Q'(I_2')}{\big(\lambda + D(r_2)K(t_2)Q(I_2')\big)^2} \partial_\gamma I_2' \big|_{\gamma=\gamma^*} + \frac{F_1(t_1)D(r_1)K(t_1)Q'(I_1')}{\big(\lambda + D(r_1)K(t_1)Q(I_1')\big)^2} \partial_\gamma I_1' \big|_{\gamma=\gamma^*} < 0 \tag{5.1.16}$$

Solving this inequality at equilibrium (5.1.14) gives

$$-\frac{D(r_2)K(t_2)Q'(I_2')}{\lambda + D(r_2)K(t_2)Q(I_2')}(-\mathcal{I}_1 S(|r_1 - r_2|) + \mathcal{I}_2) + \frac{D(r_1)K(t_1)Q'(I_1')}{\lambda + D(r_1)K(t_1)Q(I_1')}(-\mathcal{I}_1 + \mathcal{I}_2 S(|r_1 - r_2|)) < 0. \tag{5.1.17}$$

Scaling this inequality by $-F_1(t_1)F_2(t_2)$ and making use of condition (5.1.14) once more we recover the

simplified condition for ecological stability

$$F_1(t_1)D(r_2)K(t_2)Q'(I_2')(\mathcal{I}_1 S(|r_1 - r_2|) - \mathcal{I}_2) - F_2(t_2)D(r_1)K(t_1)Q'(I_1')(\mathcal{I}_1 - \mathcal{I}_2 S(|r_1 - r_2|)) < 0. \quad (5.1.18)$$

We refer to the latter directly when discussing ecological stability in the context of explicit examples. Ecological stability is showcased in Figure 5.1 for which we remark that, while in principle both quantities P_1 and P_2 are scalar functions of five variables (r_1, t_1) and (r_2, t_2) and the proportion γ the plot is showing a picture of a section of the surface $P_2 - P_1$ in the vicinity of the equilibrium proportion γ^* given through (5.1.14) for fixed traits (r_1, t_1) and (r_2, t_2). We emphasize that the plot is showing the difference in payoffs $P_2 - P_1$ in the local vicinity of γ^* to avoid confusion about the dependence on γ of $P_2 - P_1$ following a linear drop-off.

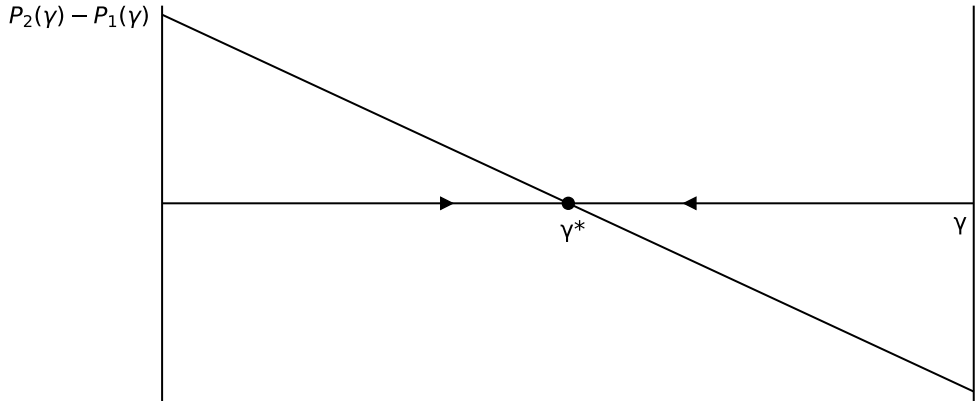

Figure 5.1: The figure illustrates condition (5.1.15). It is clear that values of $\gamma < \gamma^*$ result in the first type having smaller payoff than the second type playing (r_2, t_2), which leads to the fractional increase in the size of the latter (γ) until it reaches the critical value γ^*, at which both types grow at the same rate. If, on the other hand $\gamma > \gamma^*$, the converse is true so that γ decreases until it re-settles at γ^*, indicating that γ^* is a parameter value associated with stable co-existence.

Although it is true that ecological stability is necessary for the stable co-existence of two types over the long-run this condition is not sufficient. Indeed, we can conceive of a scenario in which the two types play (r_1, t_1) and (r_2, t_2) in a proportion that satisfies conditions (5.1.13) and (5.1.15) but in which the fitness of some mutant (say, a type-1 mutant) is larger than the resident fitness of that type, so that for some $(\hat{r}, \hat{t})$ in the local vicinity of (r_1, t_1) we have $P_1^\dagger(\hat{r}, \hat{t}; r_1, t_1; r_2, t_2) > P_1(r_1, t_1; r_2, t_2)$. In such a scenario the type-1 resident risks invasion from mutants of its type, which could in turn jeopardise that type's stable co-existence with the type-2 group over longer time-scales. In the part that follows we extend the notions of evolutionary stability as introduced in section 2 to describe evolutionary stability for a prey population consisting of two aposematic types. This is extended in a straightforward way by imposing that each type is evolutionarily stable against mutations that are local to its type.

Evolutionary stability

We say that two types can co-exist in a sense that is (locally) evolutionarily stable if each type is non-invasible by (rare) local mutations (i.e. if each type is a local ESS - see Definition 5.1.1 below). From Definition 2.3.1 it should be clear that the precise conditions for a type to be a local ESS depends on which subset $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$ or $\mathcal{D}_3$ it draws its strategy from. Since for each type there are four possibilities, it follows that one can

conceive of sixteen distinct instances in which the conditions for the system to be (locally) evolutionarily stable are unique. While studying any number of these might be of interest for general modelling purposes, we presently narrow our attention to a subclass of these that more closely resemble mimicry systems.

In order to keep the discussion both focused and general, we account for two features of mimicry systems that will help us narrow our attention. The first is that one type (type-1) is more aversive than the other (i.e. $0 \leq t_2 < t_1$). This restriction alone means that of the sixteen regimes of co-existence there are now eight that are of interest (as we are now excluding the possibility that type-1s draw their strategy from either $\mathcal{D}_0$ or $\mathcal{D}_1$). The second is that both types produce signals that are conspicuous. [3] This last restriction forces us to exclude regimes in which either type is cryptic and therefore through the earlier restriction we are now left with two regimes, since for the first type we have $(r_1, t_1) \in \mathcal{D}_3$ and for the second type we have $(r_2, t_2) \in \mathcal{D}_1$ or $(r_2, t_2) \in \mathcal{D}_3$. We hence proceed to give a formal definition of what it means for a prey population in which the first type draws its strategy (r_1, t_1) from $\mathcal{D}_1$ and the second type draws its strategy (r_2, t_2) either from $\mathcal{D}_1$ or from $\mathcal{D}_3$[4].

Definition 5.1.1. *In a prey population in which the proportion of type-1s playing (r_1, t_1) is γ and the proportion of type-2s playing (r_2, t_2) is $(1-\gamma)$ we have that:*

[i] The type-1s are locally evolutionarily stable if they receive higher fitness when interacting with the type-1 strategy than do the mutants (that are local to the type-1 strategy) when interacting with the type-1 strategy. That is, we say that the type-1s are locally evolutionarily stable if

$$P_1^{\dagger}(r = r_1, t = t_1; r_1, t_1; r_2, t_2) > P_1^{\dagger}(r, t; r_1, t_1; r_2, t_2) \text{ for all } (r, t) \in [r_1 - \delta r, r_1 + \delta r] \times [t_1 - \delta t, t_1 + \delta t] \setminus (r_1, t_1). \tag{5.1.19}$$

[ii] The type-2s are locally evolutionarily stable if they receive higher fitness when interacting with the type-2 strategy than do the mutants (that are local to the type-2 strategy) when interacting with the type-2 strategy.

(a) If the type-2s are completely undefended - i.e. $t_2 = 0$ with $(r_2, 0) \in \mathcal{D}_1$ - then they are locally evolutionarily stable if

$$P_2^{\dagger}(r = r_2, t = 0; r_1, t_1; r_2, 0) > P_2^{\dagger}(r, t; r_1, t_1; r_2, 0) \text{ for all } (r, t) \in [r_2 - \delta r, r_2 + \delta r] \times [0, \delta t] \setminus (r_2, 0). \tag{5.1.20}$$

(b) If the type-2s are defended - i.e. $0 < t_2 < t_1$ with $(r_2, t_2) \in \mathcal{D}_3$ - then they are locally evolutionarily stable if

$$P_2^{\dagger}(r = r_2, t = t_2; r_1, t_1; r_2, t_2) > P_2^{\dagger}(r, t; r_1, t_1; r_2, t_2) \text{ for all } (r, t) \in [r_2 - \delta r, r_2 + \delta r] \times [t_2 - \delta t, t_2 + \delta t] \setminus (r_2, t_2). \tag{5.1.21}$$

Finally, we say that the prey population is locally evolutionarily stable if both the type-1s and the type-2s are locally evolutionarily stable.

For the remainder of this chapter we make use of a theorem (Theorem 5.1.2) - this is a direct extension of Theorem 2.3.2 used in chapter 2 - to determine when systems involving the co-existence of two types as described in Definition 5.1.1 are locally evolutionarily stable. We state and prove this theorem below.

[3] In the worked examples of sections 5.3 and 5.4 we impose that the conspicuousness of these signals is identical.

[4] The remaining fourteen regimes of co-existence mentioned above could have been included in Definition 5.1.1, but such a task would go beyond our scope and would interrupt the flow of the manuscript.

Theorem 5.1.2. *Assume that the prey population is made up of a proportion $1-\gamma$ of type-1s playing $(r_1, t_1) \in \mathcal{D}_3$ and a proportion γ of type-2s playing $(r_2, t_2) \in \mathcal{D}_1 \sqcup \mathcal{D}_3$. Assume that the fitnesses received by mutants playing strategy (r, t) local to the type-1/2 values are denoted $P_1^\dagger / P_2^\dagger$ and given through (5.1.7)/(5.1.9). The quantities $\bar{P}_1^\dagger$ an $\bar{P}_2^\dagger$ are almost everywhere $\mathcal{C}^l$ with $l \geq 2$ except at $r = r_1$ and $r = r_2$ where they are not r-differentiable but are continuous at those values. Then the following conditions hold for determining when a resident strategy is a local ESS.*
[i] If

$$\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1} = 0 \tag{5.1.22}$$

$$\partial_{tt} P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1} < 0 \tag{5.1.23}$$

$$\overleftarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1} > 0 \ \textit{and} \tag{5.1.24}$$

$$\overrightarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1} < 0, \tag{5.1.25}$$

then the type-1s are locally evolutionarily stable.

[ii] If either

(a) $(r_2, t_2) \in \mathcal{D}_1$ (i.e. $t_2 = 0$) such that

$$\overrightarrow{\partial}_t P_2^\dagger(r, t; r_1, t_1; r_2, 0)|_{r=r_2, t=0} < 0 \tag{5.1.26}$$

$$\overleftarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, 0)|_{r=r_2, t=0} > 0 \ \textit{and} \tag{5.1.27}$$

$$\overrightarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, 0)|_{r=r_2, t=0} < 0, \tag{5.1.28}$$

or

(b) $(r_2, t_2) \in \mathcal{D}_3$ (i.e. $t_2 > 0$) such that

$$\partial_t P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2} = 0 \tag{5.1.29}$$

$$\partial_{tt} P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2} < 0 \tag{5.1.30}$$

$$\overleftarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2} > 0 \ \textit{and} \tag{5.1.31}$$

$$\overrightarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2} < 0, \tag{5.1.32}$$

then the type-2s are locally evolutionarily stable

If [i] holds together with either (a) or (b) of [ii] then the prey population is locally evolutionarily stable.

Proof. We show that inequalities (5.1.22), (5.1.23), (5.1.24) and (5.1.25) in [i] lead to local ESS for type-1 in the sense of (5.1.19) of Definition 5.1.1 by first expressing the mutant traits in terms of spherical coordinates (x, ϕ) in $\mathbb{R}^2$ so that

$$(x, \phi) \to (r, t): \quad r = r_1 + x\cos\phi \ \text{ and } \ t = t_1 + x\sin\phi \tag{5.1.33}$$

and hence expressing the type-1 mutant fitness $\mathcal{P}_1^\dagger$ in terms of the transformed coordinates x, ϕ so that

$$\mathcal{P}_1^\dagger : \mathbb{R}^{\geq 0} \times [0, 2\pi) \to \mathbb{R}^{\geq 0} : \ \ \mathcal{P}_1^\dagger(x, \phi) := P_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_2, t_2). \tag{5.1.34}$$

From the latter it is clear that the desired inequality (5.1.19) amounts to showing

$$\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) < 0 \ \text{ for all } \ \phi \in [0, 2\pi). \tag{5.1.35}$$

We proceed to showing this for cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2, 3\pi/2$, $\phi \in (\pi/2, \pi)$, $\phi = \pi$, $\phi \in (\pi, 3\pi/2)$ and $\phi \in (3\pi/2, 2\pi)$.

If $\phi = 0$ mutation occurs along the r-direction so that

$$\begin{aligned}\mathcal{P}_1^\dagger(x, \phi = 0) - \mathcal{P}_1^\dagger(0, \phi = 0) &= P_1^\dagger(r, t_1; r_1, t_1; r_2, t_2) - P_1^\dagger(r_1, t_1; r_1, t_1; r_2, t_2) \\ &\approx \underbrace{(r - r_1)}_{>0} \times \underbrace{\overrightarrow{\partial}_r \mathcal{P}_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{<0} < 0.\end{aligned} \tag{5.1.36}$$

If $\phi \in (0, \pi/2)$ we have

$$\begin{aligned}&\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_1^\dagger(x, \phi)|_{x=0, \phi\in(0,\pi/2)} \\ &= x\partial_x P_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi\in(0,\pi/2)} \\ &= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{=0} < 0.\end{aligned} \tag{5.1.37}$$

If $\phi = \pi/2$ or $\phi = 3\pi/2$ mutation is along the t-direction so that $x = t - t_1$. Furthermore, since we are assuming that the mutant fitness $\mathcal{P}_1^\dagger$ is as in (5.1.7) it follows that we can evaluate the difference in fitness between the mutant and the type-1 value using a second order expansion, so that

$$\begin{aligned}&\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) = P_1^\dagger(r_1, t; r_1, t_1; r_2, t_2) - P_1^\dagger(r_1, t_1; r_1, t_1; r_2, t_2) \\ &= (t - t_1) \times \underbrace{\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{=0} + \frac{1}{2}\underbrace{(t - t_1)^2}_{>0} \times \underbrace{\partial_{tt} P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{<0} < 0.\end{aligned} \tag{5.1.38}$$

If $\phi \in (\pi/2, \pi)$ the difference between the mutant and the type-2 fitness is given by

$$\begin{aligned}&\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_1^\dagger(x, \phi)|_{x=0, \phi\in(\pi/2,\pi)} \\ &= x\partial_x P_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi\in(\pi/2,\pi)} \\ &= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{>0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{=0} < 0.\end{aligned} \tag{5.1.39}$$

If $\phi = \pi$ we have mutation along the r-direction so that $x = r - r_1 < 0$ and thus

$$\begin{aligned}\mathcal{P}_1^\dagger(x, \phi = \pi) - \mathcal{P}_1^\dagger(0, \phi = \pi) =& P_1^\dagger(r, t_1; r_1, t_1; r_2, t_2) - P_1^\dagger(r_1, t_1; r_1, t_1; r_2, t_2) \\ =& \underbrace{(r - r_1)}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{>0} < 0. \end{aligned} \tag{5.1.40}$$

If $\phi \in (\pi, 3\pi/2)$

$$\begin{aligned}&\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_1^\dagger(x, \phi)|_{x=0, \phi \in (\pi, 3\pi/2)} \\ &= x\partial_x P_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi \in (\pi, 3\pi/2)} \\ &= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{>0} + x\sin\phi \times \underbrace{\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{=0} < 0. \end{aligned} \tag{5.1.41}$$

Finally, if $\phi \in (3\pi/2, 2\pi)$ it follows that

$$\begin{aligned}&\mathcal{P}_1^\dagger(x, \phi) - \mathcal{P}_1^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_1^\dagger(x, \phi)|_{x=0, \phi \in (3\pi/2, 2\pi)} \\ &= x\partial_x P_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi \in (3\pi/2, 2\pi)} \\ &= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{<0} + x\sin\phi \times \underbrace{\partial_t P_1^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_1, t=t_1}}_{=0} < 0. \end{aligned} \tag{5.1.42}$$

We conclude that we have shown case $[i]$ of Theorem 5.1.2 by showing that inequality (5.1.35) applies for all possible directions.

We proceed to showing case $[ii]$ of Theorem 5.1.2. As such, we will show case (a) first, and in particular that inequalities (5.1.26), (5.1.27) and (5.1.28) lead to a local ESS for the second type in the sense of (5.1.21) in Definition 5.1.1. Much like in case $[i]$, we show this by showing that

$$\mathcal{P}_2^\dagger(x, \phi) - \mathcal{P}_2^\dagger(0, \phi) < 0 \text{ for all } \phi \in [0, \pi], \tag{5.1.43}$$

where $\mathcal{P}_2^\dagger$ represents the type-2 mutant fitness expressed in terms of polar coordinates x, ϕ. That is

$$\mathcal{P}_2^\dagger : \mathbb{R}^{\geq 0} \times [0, \pi] \to \mathbb{R}^{\geq 0} : \ \mathcal{P}_2^\dagger(x, \phi) := P_2^\dagger(r = r_2 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_2, 0) \tag{5.1.44}$$

with

$$(x, \phi) \to (r, t) : \ r = r_2 + x\cos\phi \text{ and } t = x\cos\phi. \tag{5.1.45}$$

We show (5.1.43) by showing that it applies for cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2$, $\phi \in (\pi/2, \pi)$ and $\phi = \pi$ individually.

If $\phi = 0$ mutation is along the (positive) r-direction and in the vicinity of $(r_2, 0)$ so that

$$\begin{aligned}\mathcal{P}_2^\dagger(x, \phi = 0) - \mathcal{P}_2^\dagger(0, \phi = 0) =& P_2^\dagger(r, 0; r_1, t_1; r_2, 0) - P_2^\dagger(r_2, 0; r_1, t_1; r_2, 0) \\ \approx& \underbrace{(r - r_2)}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, 0)|_{r=r_2, t=0}}_{<0} < 0. \end{aligned} \tag{5.1.46}$$

If $\phi \in (0, \pi/2)$ we have

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) \approx x\partial_x \mathcal{P}_2^\dagger(x,\phi)|_{x=0,\phi\in(0,\pi/2)}$$

$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_2, 0)|_{x=0,\phi\in(0,\pi/2)}$$

$$= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_t P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{<0} < 0. \tag{5.1.47}$$

If $\phi = \pi/2$ mutation is solely along the t-direction so that

$$\mathcal{P}_2^\dagger(x,\phi=\pi/2) - \mathcal{P}_2^\dagger(0,\phi=\pi/2) = P_2^\dagger(r_2,t;r_1,t_1;r_2,0) - P_2^\dagger(r_2,0;r_1,t_1;r_2,0)$$

$$\approx t \times \underbrace{\overrightarrow{\partial}_t P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{<0} < 0. \tag{5.1.48}$$

If $\phi \in (\pi/2, \pi)$ the incremental difference between mutant and mimic fitness is negative since

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) \approx x\partial_x \mathcal{P}_2^\dagger(x,\phi)|_{x=0,\phi\in(\pi/2,\pi)}$$

$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_2, 0)|_{x=0,\phi\in(\pi/2,\pi)}$$

$$= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{>0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_t P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{<0} < 0. \tag{5.1.49}$$

To conclude part (a) of case $[ii]$ of Theorem 5.1.2 we show that for the scenario $\phi = \pi$ in which mutation is solely along the r-direction we have

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) = P_2^\dagger(r,0;r_1,t_1;r_2,0) - P_2^\dagger(r_2,0;r_1,t_1;r_2,0)$$

$$\approx \underbrace{(r-r_2)}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_2^\dagger(r,t;r_1,t_1;r_2,0)|_{r=r_2,t=0}}_{>0} < 0. \tag{5.1.50}$$

To show case (b) in part $[ii]$ of Theorem 5.1.2 it remains for us to show that equality (5.1.29) together with inequalities (5.1.30), (5.1.31) and (5.1.32) imply that the second type admits a local ESS in the sense of (5.1.21) of Definition 5.1.1. We begin by expressing the mutant traits in the vicinity of the type-2 strategy in terms of spherical coordinates x and ϕ so that

$$(x,\phi) \to (r,t): \quad r = r_2 + x\cos\phi \ \text{ and } \ t = t_2 + x\sin\phi. \tag{5.1.51}$$

In terms of the transformed coordinates x, ϕ the type-2 mutant fitness $\mathcal{P}_2^\dagger$ reads

$$\mathcal{P}_2^\dagger : \mathbb{R}^{\geq 0} \times [0, 2\pi) \to \mathbb{R}^{\geq 0}: \quad \mathcal{P}_2^\dagger(x,\phi) := P_2^\dagger(r = r_2 + x\cos\phi, t = t_2 + x\sin\phi; r_1, t_1; r_2, t_2). \tag{5.1.52}$$

From the latter it is clear that the desired inequality (5.1.21) amounts to showing

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) < 0 \ \text{ for all } \ \phi \in [0, 2\pi). \tag{5.1.53}$$

We proceed to showing this for cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2, 3\pi/2$, $\phi \in (\pi/2, \pi)$, $\phi = \pi$, $\phi \in (\pi, 3\pi/2)$

and $\phi \in (3\pi/2, 2\pi)$ individually.

If $\phi = 0$ mutation occurs along the r-direction so that

$$\mathcal{P}_2^\dagger(x, \phi = 0) - \mathcal{P}_2^\dagger(0, \phi = 0) = P_2^\dagger(r, t_2; r_1, t_1; r_2, t_2) - P_2^\dagger(r_2, t_2; r_1, t_1; r_2, t_2)$$
$$\approx \underbrace{(r - r_2)}_{>0} \times \underbrace{\overrightarrow{\partial}_r \mathcal{P}_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{<0} < 0. \tag{5.1.54}$$

If $\phi \in (0, \pi/2)$ we have

$$\mathcal{P}_2^\dagger(x, \phi) - \mathcal{P}_2^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_2^\dagger(x, \phi)|_{x=0, \phi \in (0,\pi/2)}$$
$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = t_2 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi \in (0,\pi/2)}$$
$$= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{=0} < 0. \tag{5.1.55}$$

If $\phi = \pi/2$ or $\phi = 3\pi/2$ mutation is along the t-direction so that $x = t - t_2$. Since we are assuming the mutant fitness $\mathcal{P}_2^\dagger$ is as in (5.1.9) it follows that we can evaluate the incremental differences in fitness between the mutant and the the type-2 value using an expansion with terms of up to second order. We have

$$\mathcal{P}_2^\dagger(x, \phi) - \mathcal{P}_2^\dagger(0, \phi) = P_2^\dagger(r_2, t; r_1, t_1; r_2, t_2) - P_2^\dagger(r_2, t_2; r_1, t_1; r_2, t_2)$$
$$= (t - t_2) \times \underbrace{\partial_t P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{=0} + \frac{1}{2}\underbrace{(t - t_2)^2}_{>0} \times \underbrace{\partial_{tt} P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{<0} < 0. \tag{5.1.56}$$

If $\phi \in (\pi/2, \pi)$ the difference between the mutant and the model fitness is given by

$$\mathcal{P}_2^\dagger(x, \phi) - \mathcal{P}_2^\dagger(0, \phi) \approx x\partial_x \mathcal{P}_2^\dagger(x, \phi)|_{x=0, \phi \in (\pi/2,\pi)}$$
$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = t_2 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0, \phi \in (\pi/2,\pi)}$$
$$= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{>0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{=0} < 0. \tag{5.1.57}$$

If $\phi = \pi$ we have mutation along the r-direction so that $x = r - r_2 < 0$ and thus

$$\mathcal{P}_2^\dagger(x, \phi = \pi) - \mathcal{P}_2^\dagger(0, \phi = \pi) = P_2^\dagger(r, t_2; r_1, t_1; r_2, t_2) - P_2^\dagger(r_2, t_2; r_1, t_1; r_2, t_2)$$
$$= \underbrace{(r - r_2)}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_2^\dagger(r, t; r_1, t_1; r_2, t_2)|_{r=r_2, t=t_2}}_{>0} < 0. \tag{5.1.58}$$

If $\phi \in (\pi, 3\pi/2)$ we write

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) \approx x\partial_x \mathcal{P}_2^\dagger(x,\phi)|_{x=0,\phi\in(\pi,3\pi/2)}$$

$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = t_2 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0,\phi\in(\pi,3\pi/2)}$$

$$= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r P_2^\dagger(r,t;r_1,t_1;r_2,t_2)|_{r=r_2,t=t_2}}_{>0} + x\sin\phi \times \underbrace{\partial_t P_2^\dagger(r,t;r_1,t_1;r_2,t_2)|_{r=r_2,t=t_2}}_{=0} < 0. \tag{5.1.59}$$

Finally, if $\phi \in (3\pi/2, 2\pi)$ it follows that

$$\mathcal{P}_2^\dagger(x,\phi) - \mathcal{P}_2^\dagger(0,\phi) \approx x\partial_x \mathcal{P}_2^\dagger(x,\phi)|_{x=0,\phi\in(3\pi/2,2\pi)}$$

$$= x\partial_x P_2^\dagger(r = r_2 + x\cos\phi, t = t_2 + x\sin\phi; r_1, t_1; r_2, t_2)|_{x=0,\phi\in(3\pi/2,2\pi)}$$

$$= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r P_2^\dagger(r,t;r_1,t_1;r_2,t_2)|_{r=r_2,t=t_2}}_{<0} + x\sin\phi \times \underbrace{\partial_t P_2^\dagger(r,t;r_1,t_1;r_2,t_2)|_{r=r_2,t=t_2}}_{=0} < 0, \tag{5.1.60}$$

which concludes our proof for case $[ii]$ of Theorem 5.1.2. □

We now proceed to expressing these conditions in terms of the functional forms F, D, K, Q, L, H and S, which is achieved by direct substitution of (5.1.7) into (5.1.22) through to (5.1.24) and of (5.1.9) into (5.1.26) through to (5.1.32). Careful consideration of (5.1.12) leads us to the conclusion that inequalities describing mutant invasion along r, including (5.1.24), (5.1.25) for the first type and (5.1.27), (5.1.28) or (5.1.31), (5.1.32) for the second type need be considered separately from those along the t-direction. We treat cases $r_1 < r_2$, $r_1 = r_2$ and $r_1 > r_2$ individually and in closing this section deduce that this treatment quickly allows us to rule out a number of co-existence regimes as unstable without relying on specific functional forms for the claims to hold true.

Since we have narrowed our attention to cases in which the first type is conspicuous and aversive (and therefore draws its strategy from $\mathcal{D}_3$) it follows from substitution of the fitness function (5.1.7) into (5.1.22) that

$$\frac{\lambda + D(r_1)K(t_1)Q(I_1')}{D(r_1)K(t_1)Q(I_1')}\frac{F_1'(t_1)}{F_1(t_1)} - \frac{K'(t_1)}{K(t_1)} - a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1')}{Q(I_1')} = 0. \tag{5.1.61}$$

Requiring sections of the curve in (5.1.61) to satisfy (5.1.23) amounts to condition

$$-\frac{\lambda + D(r_1)K(t_1)Q(I_1')}{D(r_1)K(t_1)Q(I_1')}\frac{F_1''(t_1)}{F_1(t_1)} + \frac{K''(t_1)}{K(t_1)} + 2a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1')}{Q(I_1')}\frac{K'(t_1)}{K(t_1)}$$

$$+a\mathcal{I}_1\frac{H''(t_1)}{H(t_1)}\frac{Q'(I_1')}{Q(I_1')} + a^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\frac{Q''(I_1')}{Q(I_1')} > 0. \tag{5.1.62}$$

As for the second type, there are two possibilities for uninvasibility along t. If it is undefended such that $(r_2, t_2) \in \mathcal{D}_1$ with $t_2 = 0$ then inequality (5.1.26) suffices to describe non-invasibility along t. Through (5.1.9) this reads

$$\frac{\lambda + D(r_2)K(0)Q(I_2')}{D(r_2)K(0)Q(I_2')}\frac{F_2'(0)}{F_2(0)} - \frac{K'(0)}{K(0)} - a\mathcal{I}_2\frac{H'(0)}{H(0)}\frac{Q'(I_2')}{Q(I_2')} < 0. \tag{5.1.63}$$

In the case that the second type is defended with $(r_2, t_2) \in \mathcal{D}_3$ it follows from (5.1.7) that (5.1.29) and

(5.1.30) amount to

$$\frac{\lambda + D(r_2)K(t_2)Q(I_2')}{D(r_2)K(t_2)Q(I_2')}\frac{F_2'(t_2)}{F_2(t_2)} - \frac{K'(t_2)}{K(t_2)} - a\mathcal{I}_2\frac{H'(t_2)}{H(t_2)}\frac{Q'(I_2')}{Q(I_2')} = 0 \tag{5.1.64}$$

and

$$-\frac{\lambda + D(r_2)K(t_2)Q(I_2')}{D(r_2)K(t_2)Q(I_2')}\frac{F''(t_2)}{F(t_2)} + \frac{K''(t_2)}{K(t_2)} + 2a\mathcal{I}_2\frac{H'(t_2)}{H(t_2)}\frac{Q'(I_2')}{Q(I_2')}\frac{K'(t_2)}{K(t_2)}$$

$$+a\mathcal{I}_2\frac{H''(t_2)}{H(t_2)}\frac{Q'(I_2')}{Q(I_2')} + a^2\mathcal{I}_2^2\left(\frac{H'(t_2)}{H(t_2)}\right)^2\frac{Q''(I_2')}{Q(I_2')} > 0. \tag{5.1.65}$$

We now consider invasion along r. If $r_1 < r_2$ then for the first type (5.1.24) and (5.1.25) read

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}\frac{L'(r_1)}{L(r_1)} - (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}S'(0) - (1-a)\gamma\mathcal{I}_2\frac{Q'(I_1')}{Q(I_1')}S'(|r_1 - r_2|) < 0 \tag{5.1.66}$$

and

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}\frac{L'(r_1)}{L(r_1)} + (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}S'(0) - (1-a)\gamma\mathcal{I}_2\frac{Q'(I_1')}{Q(I_1')}S'(|r_1 - r_2|) > 0. \tag{5.1.67}$$

Still with $r_1 < r_2$ conditions (5.1.27)/(5.1.31) and (5.1.28)/(5.1.32) for the second type read

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}\frac{L'(r_2)}{L(r_2)} + (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_2')}{Q(I_2')}S'(|r_1 - r_2|) - (1-a)\gamma\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}S'(0) < 0 \tag{5.1.68}$$

and

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}\frac{L'(r_2)}{L(r_2)} + (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_2')}{Q(I_2')}S'(|r_1 - r_2|) + (1-a)\gamma\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}S'(0) > 0, \tag{5.1.69}$$

where it is assumed that $(r_2, t_2) \in \mathcal{D}_1 \sqcup \mathcal{D}_3$.

The second case we consider has $r_2 < r_1$. For the first type (5.1.24) and (5.1.25) now read

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}\frac{L'(r_1)}{L(r_1)} - (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}S'(0) + (1-a)\gamma\mathcal{I}_2\frac{Q'(I_1')}{Q(I_1')}S'(|r_1 - r_2|) < 0, \tag{5.1.70}$$

and

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}\frac{L'(r_1)}{L(r_1)} + (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_1')}{Q(I_1')}S'(0) + (1-a)\gamma\mathcal{I}_2\frac{Q'(I_1')}{Q(I_1')}S'(|r_1 - r_2|) > 0, \tag{5.1.71}$$

respectively. For the second type (5.1.27)/(5.1.31) and (5.1.28)/(5.1.32) read

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}\frac{L'(r_2)}{L(r_2)} - (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_2')}{Q(I_2')}S'(|r_1 - r_2|) - (1-a)\gamma\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}S'(0) < 0, \tag{5.1.72}$$

and

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}\frac{L'(r_2)}{L(r_2)} - (1-a)(1-\gamma)\mathcal{I}_1\frac{Q'(I_2')}{Q(I_2')}S'(|r_1 - r_2|) + (1-a)\gamma\mathcal{I}_2\frac{Q'(I_2')}{Q(I_2')}S'(0) > 0. \tag{5.1.73}$$

The conditions detailed thus far provide a generalised extension to the single-species description of Broom et al. (2006) to account for a prey population that consists of two types. Although co-existence regimes outside of the mimetic case are not a priority there are a number of conclusions that are general and worth noting.

Examples of systems that are unstable

In Broom et al. (2006) it was demonstrated that the conspicuous signalling of non-aversive strategies is not evolutionarily stable, since such types risk invasion from their mutant counterparts that are marginally less conspicuous. Presently, we recover a related conclusion, namely that the co-existence of two attractive types is not evolutionarily stable if both of these advertise their attractiveness through conspicuous signals. Indeed, if both types are attractive it follows that $\mathcal{I}_1$ and $\mathcal{I}_2$ are negative quantities so that if, say $r_1 < r_2$ it is clear that the terms on the LHS of (5.1.66) are positive and therefore inequality (5.1.66) is violated. This suggests that a mutant type that is marginally less conspicuous exhibits a marginal fitness advantage compared with the type-1 resident, which is reasonable, since it is predicted to experience lower rates of detection (first term), to be recollected less as a type that is attractive (second term) and to benefit from its imperfect resemblance to the attractive type-1s (third term) but also to the attractive type-2s (fourth term). In particular, since the first type is not locally evolutionarily stable, the complex as a whole is not stable - see Definition 5.1.1. Notice that if we had set $r_1 > r_2$ the argument would still hold for the terms on the LHS of (5.1.72) suggesting that the type-2 risks invasion from the less conspicuous mutant type. As we see in the section that follows and as is perhaps expected, the co-existence of two types that are attractive and equally conspicuous with $r_1 = r_2 > 0$ (non-cryptic) is also not possible, since both types would risk invasion from the less conspicuous mutant.

Maintaining that $0 < r_1 < r_2$ but considering the case in which the first type is aversive and the second type is attractive such that $t_2 < t_c < t_1$ and $\mathcal{I}_1 > 0$ but $\mathcal{I}_2 < 0$ leads us to the observation that the described scenario is also not stable. Indeed, by inspection we conclude that the terms on the LHS of (5.1.68) are all positive so that (5.1.68) is violated, which suggests that the second type risks being invaded by the less conspicuous mutant. This is a sensible conclusion to draw since the less conspicuous mutant is predicted to experience a fitness advantage associated with a reduced rate of detection (first term), reduced rate of recollection as a type that is attractive (second term) a higher fitness associated with more closely resembling the aversive type-1 and a higher fitness associated with not perfectly resembling its own type, which is attractive (fourth term). Once more, we mention that these arguments would hold by inspection of (5.1.72) in the case that $r_2 < r_1$ providing that now type-1s are the attractive types, i.e. $t_1 < t_c < t_1$. We therefore conclude that an aversive and an attractive type where the latter exhibits a stronger signalling component than the former cannot co-exist in the sense of Definition 5.1.1, precisely because the second type risks being invaded by less conspicuous mutations.

As a final example, we mention that it is not possible for two aversive types (such that $\mathcal{I}_1 > 0$ and $\mathcal{I}_2 > 0$) and with with vastly different signalling components to co-exist in the sense of (5.1.6) and (5.1.15) - i.e. over the shorter (ecological) time-scale. Indeed, if we assume that either $r_1 \gg r_2$ or $r_2 \gg r_1$ such that $S(|r_1 - r_2|) \approx 0$ we recover a condition which is simplified in comparison with condition (5.1.18). This now reads

$$F_1(t_1)D(r_2)K(t_2)Q'(I_2')\mathcal{I}_2 + F_2(t_2)D(r_1)K(t_1)Q'(I_1')\mathcal{I}_1 > 0 \tag{5.1.74}$$

and does not hold. Indeed, since Q is almost everywhere monotonic decreasing it follows that $Q'(I_1') < 0$

and $Q'(I_2') < 0$ so that since $\mathcal{I}_1$ and $\mathcal{I}_2$ are both positive it follows that both terms on the LHS of (5.1.74) are negative. We have made three interesting observations: (i) the co-existence of two attractive types in which at least one has non-zero signalling component is not evolutionarily stable; (ii) the co-existence of one aversive type and one attractive type is not evolutionarily stable if the attractive type has the stronger signalling component; (iii) the co-existence of two aversive types with vastly different signalling components is not ecologically stable.

5.2 Models vs. mimics

Introducing a second type has not only doubled the conditions required for ESS analysis, but has urged us to consider a new set of conditions altogether (for ecological stability), such that if the latter do not hold there is no sense in discussing the former. One could rightly argue that there is no value in determining the evolutionary stability of a complex over the prolonged (evolutionary) time-scale if this is not stable on the shorter time-scale: a pair of strategies that is evolutionarily stable when played in specific proportions may not be ecologically stable in those proportions so that the composition is likely to change within a timescale that is shorter than the timescale under which we would consider its evolutionary stability. In keeping with the mimicry focus of this chapter, we presently identify the first type with the aversive model and the second type with a completely undefended mimic who perfectly resembles the model; that is, we set $r_1 = r_2$ and $0 = t_2 < t_c < t_1$. While it could be argued that the assumptions of perfect mimetic resemblance and the exclusion of the possibility that the mimic is even slightly defended such that $0 < t_2 < t_c$ are restrictive we contend that they are biologically plausible and a constitute a meaningful starting point for the analysis.

Setting $r_1 = r_2$ suggests that $S(|r_1 - r_2|) = 1$ so that the expressions for perceived aversivess of either type as given in (5.1.3) and (5.1.4) amount to

$$I_1' = \big[a + (1-a)(1-\gamma)\big]\mathcal{I}_1 + (1-a)\gamma\mathcal{I}_2 \tag{5.2.1}$$

and

$$I_2' = (1-a)(1-\gamma)\mathcal{I}_1 + \big[a + (1-a)\gamma\big]\mathcal{I}_2. \tag{5.2.2}$$

So that while from the modelling perspective the described mimicry setup is notably simpler than when compared with the more general class of co-existence problems the presence of two types remains a challenge.

Eco-evolutionary stability & special remarks about the functional forms

In light of this complexity we restrict our attention to a special class of examples in which the rate of background mortality is taken to be zero (predation is the only source of prey death) so that $\lambda = 0$ and that predators have perfect recollection of encounters with prey. We remark that the latter has been integral in all worked examples considered thus far (we also use it in chapter 6) and amounts to setting $L(r) = D(r)$ so that there is one less form to consider. Recall that to model mimicry we set $r_1 = r_2$ and assume that type-1s are aversive and that type-2s are completely undefended so that $t_1 > t_c$ and $t_2 = 0$. Equal conspicuousness suggests that $S(|r_1 - r_2|) = 1$ and together with the assumption that $\lambda = 0$ the condition for ecological equilibrium given in (5.1.14) amounts to

$$\frac{Q(I_1')}{Q(I_2')} = \frac{F_1(t_1)}{K(t_1)}\frac{K(t_2)}{F_2(t_2)}. \tag{5.2.3}$$

It is also straightforward to demonstrate that the condition for ecological stability given in (5.1.18) now simplifies to

$$F_2(t_2)K(t_1)Q'(I_1') - F_1(t_1)K(t_2)Q'(I_2') > 0, \tag{5.2.4}$$

which when solved simultaneously with (5.2.3) results in a restriction on the Q functions only

$$-\frac{Q'(I_2')}{Q(I_2')} > -\frac{Q'(I_1')}{Q(I_1')}. \tag{5.2.5}$$

Since we identify type-1s with the aversive models and type-2s with the attractive mimics it follows from (5.2.1) and (5.2.2) that we have $I_2' < I_1'$ and therefore this condition suggests that in order for a given equilibrium solution to be stable, the Q function should be chosen such that $-Q'/Q$ is a positive quantity that decreases with increasing values of I'. Graphically, this means that the slope of the line tangent to a point along the graph of Q over the value of the function at that point is a negative and ever-decreasing (in absolute value) quantity. That is, ecological stability for a model-mimic complex in the $\lambda = 0$ regime requires us to consider forms for Q whose slope shrinks faster than the value of the function itself. This suggests that functions that would otherwise be natural candidates to consider such as $Q(I') = max(1, q_0 \exp(-qI'))$ shall not be considered.

Indeed, while the form $Q(I') = max(1, q_0 \exp(-qI'))$ is used extensively throughout, the specific restrictions presented in this chapter render it less suitable for considering mimicry (although it is used in chapter 6 where the specific assumptions relating to mimicry are different). Indeed, for such a form $-Q'(I')/Q(I') = -q$ for all $I' > 0$, which violates (5.2.5). Indeed, as it happens exponential forms for Q with powers greater than or equal to one violate (5.2.5) - for instance $Q(I') = \exp(-qI'^2)$ considered on $I' > 0$ give $Q'/Q = -2qI'$, which satisfies the reverse of (5.2.5). For this reason, we restrict our attention to exponential forms with power less than unity. A natural form to consider on $I' > 0$ is $Q(I') = q_0 \exp(-q\sqrt{I'})$, which clearly satisfies (5.2.5). We return to this point once we discuss evolutionary stability in the context of mimicry.

For the model to be a local ESS in the sense of (5.1.19) it must satisfy conditions (5.1.22) through to (5.1.25). In the previous section it was shown that through substitution of (5.1.7) with $= 0$ into $P_1^\dagger$ (5.1.22) and (5.1.23) amount to (5.1.61) and (5.1.62). Setting $r_1 = r_2$ and $0 = t_2 < t_c < t_1$ to model mimicry we now have for the model that

$$\frac{F_1'(t_1)}{F_1(t_1)} - \frac{K'(t_1)}{K(t_1)} - a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1')}{Q(I_1')} = 0. \tag{5.2.6}$$

and

$$-\frac{F''(t_1)}{F(t_1)} + \frac{K''(t_1)}{K(t_1)} + 2a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{Q'(I_1')}{Q(I_1')}\frac{K'(t_1)}{K(t_1)} + a^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\frac{Q''(I_1')}{Q(I_1')} > 0, \tag{5.2.7}$$

where we have assumed a linear form for H as in (3.2.1) and (4.1.1) so that the term involving H'' in (5.1.62) is taken to be vanishing.

Invasion along the t-direction is more straightforward for the mimic and is guaranteed through the single inequality in (5.1.63). Setting $\lambda = 0$, $r_1 = r_2$ and $0 = t_1 < t_c < t_1$ into this gives

$$\frac{F_2'(0)}{F_2(0)} - \frac{K'(0)}{K(0)} - a\mathcal{I}_2\frac{H'(0)}{H(0)}\frac{Q'(I_2')}{Q(I_2')} < 0. \tag{5.2.8}$$

In addition to (5.2.6) and (5.2.7) the model must satisfy inequalities (5.1.66) and (5.1.67) so as to be a

local ESS in the sense of Theorem 5.1.2. When setting $\lambda = 0$, $r_1 = r_2$ and $0 = t_2 < t_c < t_1$ these amount to

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1 \frac{Q'(I_1')}{Q(I_1')}\frac{D'(r_1)}{D(r_1)} - (1-a)(1-\gamma)\mathcal{I}_1 \frac{Q'(I_1')}{Q(I_1')}S'(0) - (1-a)\gamma\mathcal{I}_2 \frac{Q'(I_1')}{Q(I_1')}S'(0) < 0, \tag{5.2.9}$$

and

$$\frac{D'(r_1)}{D(r_1)} + a\mathcal{I}_1 \frac{Q'(I_1')}{Q(I_1')}\frac{D'(r_1)}{D(r_1)} + (1-a)(1-\gamma)\mathcal{I}_1 \frac{Q'(I_1')}{Q(I_1')}S'(0) + (1-a)\gamma\mathcal{I}_2 \frac{Q'(I_1')}{Q(I_1')}S'(0) > 0. \tag{5.2.10}$$

As for the mimic, in addition to (5.2.8) imposing local ESS through Theorem 5.1.2 suggests that (5.1.68) and (5.1.69) must also hold. Considering we are working in the mimetic regime, where $\lambda = 0$, $r_1 = r_2$ and $0 = t_1 < t_c < t_2$ these amount to

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2 \frac{Q'(I_2')}{Q(I_2')}\frac{D'(r_2)}{D(r_2)} - (1-a)(1-\gamma)\mathcal{I}_1 \frac{Q'(I_2')}{Q(I_2')}S'(0) - (1-a)\gamma\mathcal{I}_2 \frac{Q'(I_2')}{Q(I_2')}S'(0) < 0, \tag{5.2.11}$$

$$\frac{D'(r_2)}{D(r_2)} + a\mathcal{I}_2 \frac{Q'(I_2')}{Q(I_2')}\frac{D'(r_2)}{D(r_2)} + (1-a)(1-\gamma)\mathcal{I}_1 \frac{Q'(I_2')}{Q(I_2')}S'(0) + (1-a)\gamma\mathcal{I}_2 \frac{Q'(I_2')}{Q(I_2')}S'(0) > 0 \tag{5.2.12}$$

From the above it is immediately clear that a complex consisting of two attractive types with non-zero signalling component is not evolutionarily stable. Indeed, inspection of conditions (5.2.9) and (5.2.11) leads us to the conclusion that these consist of positive terms only. Based on the assumptions we have made thus far this is a reasonable conclusion to make since less conspicuous mutations occurring in either type threaten to invade on account of reduced rate of predation, reduced rate of recollection attractive types and a fitness advantage associated with imperfectly resembling a complex that is collectively perceived as attractive by the average predator. It is typical of mimicry complexes to consist of a defended type and an undefended type and it is interesting to uncover that under the present description this is the only sensible choice to consider.

In keeping with the simplifying restriction $\lambda = 0$, we restrict our attention to Q functions of an exponentially-decaying form on the domain $I' > 0$ that satisfy (5.2.5) and therefore require that the power on the exponent is less than one. We have developed a framework within which to model the co-existence of two types that are inherently different with respect to their degree of adaptation to the defence, since this would be a natural feature of real mimicry complexes in which one undefended species resembles a defended one. However, as a preliminary example we may wish to consider a more simple scenario in which the types are indistinguishable by means of their inherent properties so that any differences are solely attributed to differences in the strategies they play and the proportions in which they do so. We begin by imposing $F_1 = F_2$ and incorporate increasing layers of complexity over the sections that follow. Given the discussions about the Q function that have been brought up thus far, we remark that the most natural example to consider would be

$$F_1(t) = F_2(t) := f_0 \exp(-ft); \quad H(t) := t - t_c; \quad K(t) := \frac{k_0}{1+kt};$$

$$L(r) = D(r); \quad Q(I) := q_0 \exp\left(-q\sqrt{I}\right); \quad S(x) = max(1 - vx, 0). \tag{5.2.13}$$

Upon inspection, however, we observe that this is not possible, since substitution of (5.2.13) into the type-1

equilibrium condition (5.2.6) amounts to the equality

$$-f + \frac{k}{1+kt_1} + aq\frac{N}{2n}D(r_1)\frac{1}{\sqrt{I_1'}} = 0. \tag{5.2.14}$$

We remark that this cannot be solved in tandem with condition (5.2.8) for the second type, since on account of (5.2.13) condition (5.2.8) reads

$$-f + k + aq\frac{N}{2n}D(r_1)\frac{1}{\sqrt{I_2'}} < 0, \tag{5.2.15}$$

which together with (5.2.14) leads to the contradicting inequality

$$\sqrt{I_2'} > \frac{aq\frac{N}{2n}D(r_1)}{f-k} > \frac{aq\frac{N}{2n}D(r_1)}{f-\frac{k}{1+kt_1}} = \sqrt{I_1'}. \tag{5.2.16}$$

As it is not possible to solve the example of (5.2.13) we provide a modification to the form of K to demonstrate that even in the simplest case in which $F_1 = F_2$ we can at least obtain a stable point solution. Following this in section 5.4 we show that if we relax the condition $F_1 = F_2$ the mentioned modification in K is not required and demonstrate how we can recover a stable continuum of solutions.

5.3 A stable point solution

We closed the last section by demonstrating that the seemingly simplest regime of mimicry in which there are no intrinsic differences between the types (i.e. $F_1 = F_2$) comes at a price so that more careful consideration of the form for K is required. In this section we provide detailed numerical analysis through which we demonstrate how the example in (5.3.1), in which the modifications in K are made can be used to generate a stable *point solution*. We clarify that point solutions (to be contrasted with a continuum of solutions) consist of a specific choice of strategy for the model and the mimic and a proportion for these to be played in that satisfies the conditions for evolutionary and ecological stability.

We consider the following set of example functions

$$F_1(t) = F_2(t) := f_0\exp(-ft); \quad H(t) := t - t_c; \quad K(t) := \frac{k_0}{1+kt^2};$$

$$L(r) = D(r); \quad Q(I) := q_0\exp(-q\sqrt{I}); \quad S(x) = max(1-vx, 0). \tag{5.3.1}$$

Notice that since $F_1 = F_2$ we do not distinguish between the forms of the type-1 and type-2 mutant payoffs. Partial differentiation of the mutant payoff t-derivative of $P_1^\dagger/P_2^\dagger$ given in (5.1.7)/(5.1.9) gives

$$C(r,t)\times\partial_t P^\dagger = -f + \frac{2kt}{1+kt^2} + aq\frac{N}{2n}D(r)\frac{1}{\sqrt{I'}}, \tag{5.3.2}$$

where $C(r,t) := D(r)K(t)Q(I')/F(t)$. Notice that $K(t)$ has a maximum at $t = 1/\sqrt{k}$, whose value is $K(t = 1/\sqrt{k}) = \sqrt{k}$ and minimum of zero at $t = 0$. Setting $t_1 = 1/\sqrt{k}$ and $t_2 = 0$ maximises the contribution of the second term on the RHS of (5.3.2) for the model and minimises the contribution of this term for the mimic. We emphasise this fact merely to demonstrate that the third term in (5.3.2) is higher for the mimic than it is for the model (notice that Q satisfies (5.2.5) and the mimic is perceived as less aversive) and is crucial in resolving the issue encountered with the forms in (5.2.13) when attempting to satisfy (5.3.3)

and (5.3.4) jointly.

Indeed, setting $t_1 = 1/\sqrt{k}$ and $t_2 = 0$ into (5.3.2) we obtain the condition for the first type to be in equilibrium along the t as

$$-f + \sqrt{k} + aq\frac{N}{2n\sqrt{I_1'}}D(r_1) = 0 \tag{5.3.3}$$

and from (5.2.7) we recover

$$-f + aq\frac{N}{2n\sqrt{I_2'}}D(r_1) < 0, \tag{5.3.4}$$

where I_1' and I_2' are provided as in (5.2.1) and (5.2.2). Notice that the condition for ecological equilibrium in (5.2.3) now reads

$$\sqrt{I_2'} = \sqrt{I_1'} - \frac{f\sqrt{k}\log 2}{q\sqrt{k}}. \tag{5.3.5}$$

From comparing (5.2.1) and (5.2.2) it also holds that

$$I_2' = I_1' - a\frac{N}{n\sqrt{k}}D(r) \tag{5.3.6}$$

and indeed that

$$I_1' = \frac{N}{n}D(r)\left[\left(\frac{1}{\sqrt{k}} - t_c\right) - \gamma(1-a)\frac{1}{\sqrt{k}}\right]. \tag{5.3.7}$$

Since we have restricted $t_1 = 1/\sqrt{k}$ and $t_2 = 0$ we seek to find an optimal proportion of mimics γ and an optimal level for the common signalling of the complex r_1. We proceed as follows, noting from (5.3.3) that $f > \sqrt{k}$. Re-arranging (5.3.3) gives

$$\sqrt{I_1'} = \frac{aq\frac{N}{n}D(r)}{2(f-\sqrt{k})} \tag{5.3.8}$$

and squaring (5.3.5) gives

$$I_2' = I_1' - 2\sqrt{I_1'}\frac{\left(f - \sqrt{k}\log 2\right)}{q\sqrt{k}} + \frac{\left(f-\sqrt{k}\log 2\right)^2}{q^2k}. \tag{5.3.9}$$

From (5.3.6) and the re-arrangement (5.3.8) the latter amounts to

$$a\frac{N}{\sqrt{k}n}D(r)\left(\frac{f-\sqrt{k}\log 2}{f-\sqrt{k}} - 1\right) = \frac{(f-\sqrt{k}\log 2)^2}{q^2k} \tag{5.3.10}$$

and can be solved for $D(r_1)$ to yield the following implicit condition on the optimal value of r_1 through

$$D(r) = \frac{1}{1-\log 2}\frac{(f-\sqrt{k})(f-\sqrt{k}\log 2)^2}{aq^2k\frac{N}{n}}. \tag{5.3.11}$$

The precise optimal level of conspicuousness depends on the specific choice of D, but can be realised as the pre-image of the RHS in (5.3.11). The optimal value of γ can be solved for by looking at (5.3.7) and (5.3.3) in conjunction, such that

$$\frac{a^2q^2ND(r)}{4n(f-\sqrt{k})^2} = \frac{1-\sqrt{k}t_c}{\sqrt{k}} - \gamma\frac{(1-a)}{\sqrt{k}}. \tag{5.3.12}$$

Appropriate re-arrangement and substitution of (5.3.11) reduces the latter to

$$\gamma = \frac{1-\sqrt{k}t_c}{1-a} - \frac{1}{4-4\log 2}\frac{a}{1-a}\frac{(f-\sqrt{k\log 2})^2}{\sqrt{k}(f-\sqrt{k})}. \tag{5.3.13}$$

The recovered values for r_1, t_1 and γ must now be reviewed in tandem with the bound provided in (5.3.4), which is

$$\frac{1}{\sqrt{I_2'}} < \frac{2f}{aq\frac{N}{n}D(r_1)}, \tag{5.3.14}$$

and may be solved in tandem with (5.3.5) and (5.3.3) to give

$$aq\frac{N}{2n}D(r_1)\left(\frac{1}{f-\sqrt{k}} - \frac{1}{f}\right) > \frac{f-\sqrt{k}\log 2}{q\sqrt{k}}. \tag{5.3.15}$$

Substitution of (5.3.11) for $D(r_1)$ simplifies the latter substantially so that we recover a lower bound on f

$$\frac{f-\sqrt{k}\log 2}{2f(1-\log 2)} > 1 \quad \Leftrightarrow \quad f > \frac{\log 2}{2\log 2 - 1}\sqrt{k}. \tag{5.3.16}$$

We now show that condition (5.2.7) for the first type holds. From (5.3.1) and setting $t_1 = 1/\sqrt{k}$ gives

$$-f^2 + k + aq\frac{N}{n}D(r_1)\sqrt{k} \times \frac{1}{\sqrt{I_1'}} + \frac{1}{4} \times a^2q^2\frac{N^2}{n^2}D^2(r_1)\left(\frac{1}{q\sqrt{I_1'}} + 1\right) > 0 \tag{5.3.17}$$

Substitution of (5.3.3) and (5.3.11) gives rise to multiple term cancellations so that the latter amounts to the following, trivially true inequality

$$2k(f-\sqrt{k})^2 > 0. \tag{5.3.18}$$

It remains for us to consider invasion along r through inequalities (5.2.9) and (5.2.10) for the first type and (5.2.11), (5.2.12) for the second type. As it happens, all four inequalities can be satisfied providing v in the form for $S(x) = \max(1-vx, x)$ in (5.3.1) is chosen to be large enough.

Substitution of functional forms (5.3.1) into (5.2.9) and (5.2.10) give the required conditions for the model

$$-\frac{D'(r_1)}{D(r_1)} + aq\frac{N}{n}D'(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_1'}} + (1-a)(1-\gamma)qv\frac{N}{n}D(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_1'}}$$
$$-(1-a)\gamma qv\frac{N}{n}D(r_1)t_c \times \frac{1}{2\sqrt{I_1'}} > 0 \tag{5.3.19}$$

and

$$-\frac{D'(r_1)}{D(r_1)} + aq\frac{N}{n}D'(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_1'}} - (1-a)(1-\gamma)qv\frac{N}{n}D(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_1'}}$$
$$+(1-a)\gamma qv\frac{N}{n}D(r_1)t_c \times \frac{1}{2\sqrt{I_1'}} < 0. \tag{5.3.20}$$

Substitution of (5.3.1) into (5.2.11) and (5.2.12) gives

$$-\frac{D'(r_1)}{D(r_1)} - aq\frac{N}{n}D'(r)t_c \times \frac{1}{2\sqrt{I_2'}} + (1-a)(1-\gamma)qv\frac{N}{n}D(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_2'}}$$

$$-(1-a)\gamma qv\frac{N}{n}D(r_1)t_c \times \frac{1}{2\sqrt{I_2'}} > 0, \tag{5.3.21}$$

and

$$-\frac{D'(r_1)}{D(r_1)} - aq\frac{N}{n}D'(r_1)t_c \times \frac{1}{2\sqrt{I_2'}} - (1-a)(1-\gamma)qv\frac{N}{n}D(r_1)\left(\frac{1}{\sqrt{k}} - t_c\right) \times \frac{1}{2\sqrt{I_2'}}$$

$$+(1-a)\gamma qv\frac{N}{n}D(r_1)t_c \times \frac{1}{2\sqrt{I_2'}} < 0 \tag{5.3.22}$$

respectively. It remains for us to identify a set of numerical values that is consistent with the conditions for eco-evolutionary stability described above. We provide evidence of all working in the hope of demonstrating that the specific values chosen do not consist of a choice that is non-generic. For instance, we have assumed that the model plays $t_1 = 1/\sqrt{k}$ and is aversive so that

$$t_1 > t_c \;\; \Leftrightarrow \;\; t_c\sqrt{k} < 1. \tag{5.3.23}$$

As shown in the discussion of condition (5.3.4) - for given choice of k there is a lower bound on the value of f provided by

$$f > \frac{\log 2}{2\log 2 - 1}\sqrt{k} \approx 1.8\sqrt{k}. \tag{5.3.24}$$

For example, setting $f = 2\sqrt{k}$ would be a consistent restriction on the parameters. Before proceeding with the re-reading of the above as restrictions on the parameter space we note the following.

Implicit in this section has been the assumption that $I_2' > 0$, which suggests that the complex is collectively perceived as aversive despite the presence of a type that is attractive. This is a reasonable assumption to make, especially on account of our previous remark that the co-existence of two attractive types that resemble each other is unstable in this context. Furthermore, it is achievable providing the proportion of mimics is not too large. On account of (5.2.2) the restriction $I_2' > 0$ amounts to

$$I_2' > 0 \;\; \Leftrightarrow \;\; (1-a)(1-\gamma)(1-\sqrt{k}t_c) - a\sqrt{k}t_c - (1-a)\gamma\sqrt{k}t_c > 0, \tag{5.3.25}$$

which is equivalent to the following upper bound on the mimic proportion

$$\gamma < 1 - \frac{\sqrt{k}t_c}{1-a} =: \gamma_{MAX}. \tag{5.3.26}$$

That is, the results established thus far are subject to the above condition, and indeed, for purposes of consistency one should check that the value of γ from (5.3.13) really is below the bound γ_{MAX}. In particular, we wish to identify the values that parameter a can assume such that the following inequality holds

$$0 < \gamma < \gamma_{MAX} < 1. \tag{5.3.27}$$

First, recall that

$$\gamma = \frac{1-\sqrt{k}t_c}{1-a} - \frac{a}{1-a}\frac{1}{4-4\log 2}\frac{(f-\sqrt{k}\log 2)^2}{\sqrt{k}(f-\sqrt{k})} \tag{5.3.28}$$

and note that from the definition of of γ_{MAX} in (5.3.26) it is clear that $\gamma_{MAX} < 1$. To guarantee that the bound on γ is a sensible one we must impose that

$$\gamma_{MAX} > 0 \;\;\Leftrightarrow\;\; a < 1-\sqrt{k}t_c =: a_1. \tag{5.3.29}$$

Continuing this argument, in order for γ in (5.3.28) to admit an accurate description of the mimic fraction it must be that

$$\gamma > 0 \quad \Leftrightarrow \quad a < \frac{4(1-\log 2)\sqrt{k}(f-\sqrt{k})(1-\sqrt{k}t_c)}{(f-\sqrt{k}\log 2)^2} =: a_2 \tag{5.3.30}$$

As is expected, the upper bound on a from (5.3.30) is smaller than the upper bound of (5.3.29). By comparing these bounds it is clear that this is the case providing that

$$(f-\sqrt{k}\log 2)^2 - 4(1-\log 2)\sqrt{k}(f-\sqrt{k}) > 0,$$

or equivalently that the following concave-up (notice that the leading coefficient is positive) polynomial in $\sqrt{k}$ is strictly positive

$$(4+\log^2 2 - 4\log 2)k - 2f\sqrt{k}(2-\log 2) + f^2 > 0. \tag{5.3.31}$$

Since the discriminant of this polynomial is zero, it has a double root, which occurs at

$$f = \frac{4+\log^2 2 - 4\log 2}{2-\log 2}\sqrt{k},$$

but this root is smaller than the feasible solutions for f in (5.3.24) and therefore the latter inequality holds true. In particular, we have established that $a_2 < a_1$ so that if $\gamma > 0$ then $\gamma_{MAX} > 0$, as expected. As a final remark, we demonstrate that the bound γ_{MAX} is a true upper bound in the sense that

$$\gamma_{MAX} - \gamma > 0.$$

Indeed, substitution of γ and γ_{MAX} from (5.3.28) and (5.3.26) and through scaling by $(1-a)/a$ the latter amounts to

$$\frac{1}{4-4\log 2}\frac{(f-\sqrt{k}\log 2)^2}{\sqrt{k}(f-\sqrt{k})} - 1 > 0.$$

With appropriate re-arrangement the above amounts to

$$(4+\log^2 2 - 4\log 2)k - 2f\sqrt{k}(2-\log 2) + f^2 > 0,$$

which is the same as (5.3.31) and which was subsequently demonstrated to hold true for admissible choices of f and k. Indeed, the latter demonstrates that if $\gamma > 0$ then γ_{MAX} really does act as an upper bound such that $\gamma < \gamma_{MAX}$. If this bound condition were violated (i.e. there were some positive value of $\gamma < 1$ such that $\gamma > \gamma_{MAX}$) then this would mean that $I_2' < 0$ for some admissible combination of the parameters. Clearly, this would give rise to a contradiction since the very conditions (see (5.3.3) through to (5.3.7)) that the parameters are restricted to satisfy were derived on the assumption that $I_2' > 0$, or rather, that $\gamma < \gamma_{MAX}$.

In summary, feasible solutions to γ are guaranteed by choosing a such that

$$a < \frac{4(1-\log 2)\sqrt{k}(f-\sqrt{k})(1-\sqrt{k}t_c)}{(f-\sqrt{k}\log 2)^2}. \tag{5.3.32}$$

From (5.3.11) it is clear that this amounts to choosing q and the ratio N/n such that

$$q^2\frac{N}{n} > \frac{(f-\sqrt{k})(f-\sqrt{k}\log 2)^2}{ak(1-\log 2)}. \tag{5.3.33}$$

Finally, we are left to identify an appropriate bound on v that will solve (5.3.19), (5.3.20) and(5.3.21), (5.3.22). Working pair-wise with (5.3.19) and (5.3.20), we deduce that these can be satisfied providing

$$v > \frac{\left|aq\frac{N}{n}D'(r)(1-\sqrt{k}t_c) - 2\sqrt{k}\sqrt{I_1'}\frac{D'(r)}{D(r)}\right|}{aq\frac{N}{n}D(r)(1-\sqrt{k}t_c-\gamma)} =: v_1. \tag{5.3.34}$$

Similarly, working in pairs with (5.3.21) and (5.3.22) for the mimic type we deduce that these can be solved provided that

$$v > \frac{aq\frac{N}{n}D'(r)\sqrt{k}t_c + 2\sqrt{k}\sqrt{I_2'}\frac{D'(r)}{D(r)}}{(1-a)q\frac{N}{n}D(r)(1-\sqrt{k}t_c-\gamma)} =: v_2. \tag{5.3.35}$$

In particular, we require that both (5.3.34) and (5.3.35) are satisfied, such that

$$v > \max\big(v_1, v_2\big). \tag{5.3.36}$$

Now that an appropriate lower bound has been identified for v, we can proceed by implementing specific numerical examples. We begin by setting $k = 1$. With regards to (5.3.23) we set $t_c = 0.2 < 1$. As mentioned earlier (see (5.3.24)) an appropriate restriction on the parameters is $f = 2\sqrt{k} = 2 < 2\log 2/(2-\log 2)$. We can estimate the bound on the parameter a from (5.3.30), which reads

$$a < \frac{4(1-\log 2)\sqrt{1}(2-\sqrt{1})(1-\sqrt{1}\times 0.2)}{(2-\sqrt{1}\log 2)^2} = 3.2\frac{(1-\log 2)}{(2-\log 2)^2} \approx 0.57. \tag{5.3.37}$$

For convenience we can set $a = 0.5$. Now that parameters k, f, t_c and a have been specified, we are in position to determine γ through (5.3.28), which reads

$$\gamma = \frac{8}{5} - \frac{1}{4}\frac{(2-\log 2)^2}{1-\log 2} \approx 0.21. \tag{5.3.38}$$

Notice that, as expected this value is below the bound γ_{MAX} defined in (5.3.26), namely

$$\gamma \approx 0.21 < \gamma_{MAX} = 1 - \frac{1\times 0.2\sqrt{1}}{0.5} = 0.6. \tag{5.3.39}$$

We proceed by approximating the lower bound on the product q^2n/n through (5.3.33) by imposing that the rate of detection of prey does not exceed unity such that

$$q^2\frac{N}{n} > \frac{(2-\sqrt{1})(2-\sqrt{1}\log 2)^2}{0.5\times 1(1-\log 2)} = \frac{2(2-\log 2)^2}{1-\log 2} \approx 11.13. \tag{5.3.40}$$

For example, we may set $q = 2$ and the fraction N/n equal to 3, such the product $q^2 \times N/n$ is 12. Naturally, the optimal level of conspicuousness is defined through the D-pre-image

$$r_1 = D^{-1}\left(\frac{(2-\log 2)^2}{6(1-\log 2)}\right). \tag{5.3.41}$$

If we set the base-line rate of detection as $d_0 = 0.5$ with

$$D(r_1) = \frac{d_0}{d_0 - (1-d_0)\exp(-r_1)} = \frac{1}{1+\exp(-r_1)} \tag{5.3.42}$$

then from (5.3.41) we have

$$r_1 = \log\left(\frac{(2-\log 2)^2}{6(1-\log 2) - 2(2-\log 2)^2} \approx \right) \approx 2.55 \tag{5.3.43}$$

so that the rate of detection now amounts to

$$D(r_1) = \frac{(2-\log 2)^2}{6(1-\log 2)} \approx 0.93. \tag{5.3.44}$$

The final step that remains is for us to specify a numerical value for v and from (5.3.36). It is clear that knowledge of v_1 and v_2 is a pre-requisite. Substitution of the parameter values into (5.3.34) implies that

$$v_1 = \frac{1.2 \times D'(r_1) - 2\sqrt{I_1'} \times \frac{D'(r)}{D(r_1)}}{3D(r_1) \times (0.8-\gamma)}, \tag{5.3.45}$$

where the terms $D(r)$ and γ have precise analytical expressions as provided in (5.3.44) and (5.3.38). Indeed, the denominator of the fraction in (5.3.45) reads

$$3D(r_1) \times (0.8-\gamma) = \frac{(2-\log 2)^2}{1-\log 2} \times \left(\frac{1}{8} \times \frac{(2-\log 2)^2}{1-\log 2} - \frac{2}{5}\right) \approx 1.64, \tag{5.3.46}$$

while for the numerator of the fraction in (5.3.45) we must first evaluate $D'(r_1)$, $D'(r_1)/D(r_1)$ and $\sqrt{I_1'}$. We have

$$D'(r_1) = \frac{\exp(-r_1)}{(1+\exp(-r_1))^2} = \left[6(1-\log 2) - (2-\log 2)^2\right] \times \frac{(2-\log 2)^2}{6^2(1-\log 2)^2} \approx 0.067, \tag{5.3.47}$$

where for the latter we have made use of (5.3.43). In addition, it is true that

$$\frac{D'(r_1)}{D(r_1)} = \frac{6(1-\log 2) - (2-\log 2)^2}{6(1-\log 2)} \approx 0.072 \tag{5.3.48}$$

and using the exact forms for $D(r_1)$ and γ - see (5.3.44) and (5.3.38) - we have

$$\sqrt{I_1'} = \frac{(2-\log 2)^2}{4(1-\log 2)} \approx 1.39. \tag{5.3.49}$$

Substitution of (5.3.47), (5.3.48) and (5.3.49) implies that the numerator of the fraction in (5.3.45) is ap-

proximately

$$1.2 \times D'(r) - 2\sqrt{I_1'} \times \frac{D'(r)}{D(r)} \approx 3.14, \tag{5.3.50}$$

so that

$$v_1 \approx 1.91. \tag{5.3.51}$$

In order to determine v_2 we proceed as follows. From (5.3.35) it is evident that the denominator in v_1 is also common to v_2 and given as in (5.3.46). The numerator in v_2 reads

$$aq\frac{N}{n}D'(r_1)\sqrt{k}t_c + 2\sqrt{k}\sqrt{I_2'} = 2\sqrt{I_2'} \times \frac{D'(r_1)}{D(r_1)} + 0.6 \times D'(r_1), \tag{5.3.52}$$

where the exact numerical expressions for $D'(r)$ and $D'(r)/D(r)$ are provided in (5.3.47) and (5.3.48). The missing term therefore is $\sqrt{I_2'}$ and can be assigned an exact numerical expression using (5.3.44) and (5.3.38), such that

$$\sqrt{I_2'} = \frac{\log 2}{4} \times \frac{2 - \log 2}{1 - \log 2} \approx 0.74. \tag{5.3.53}$$

The numerator in v_2 can thus be approximated as

$$2\sqrt{I_2'} \times \frac{D'(r_1)}{D(r_1)} + 0.6 \times D'(r_1) \approx 0.15, \tag{5.3.54}$$

and so the entire fraction v_2 amounts to

$$v_2 \approx 0.089 < v_1, \tag{5.3.55}$$

so that through (5.3.36) we require that

$$v > max(v_1, v_2) \approx 1.91, \tag{5.3.56}$$

and so setting $v = 2$ solves (5.3.19)/(5.3.20) and (5.3.21)/(5.3.22) concludes the present numerical analysis. In this, we have demonstrated that if parameters $f, k, t_c, a, q, N/n$ and v chosen such that $f = 2, k = 1, t_c = 0.2, a = 0.5, q = 2, N/n = 3$ and $v = 2$ a complex consisting of a model playing $(r_1, t_1) \approx (2.55, 1)$ and a mimic playing $(r_2, t_2) \approx (2.55, 0)$ where the latter resides in proportion $\gamma \approx 0.21$ is eco-evolutionarily stable in the sense of short-term ecological stability and longer scale evolutionary stability seen in Definition 5.1.1 and Theorem 5.1.2. We argue that on account of the manner in which specific numerical values were picked the point solution $(r_1, t_1) \approx (2.55, 1)$ and $(r_2, t_2) \approx (2.55, 0)$ is part of a continuum of solutions, whose exact extent over the strategy space we do not discuss presently. Nonetheless, the explicit workings provided in this section should convince the reader that we have not come across an example that is non-generic.

Mimicry as discussed in this section is conceivably the most straightforward adaptation of the single-type description provided in chapter 2 and serves as a preliminary introduction to the mathematical modelling of mimicry. In this, we have treated mimics as a group of undefended individuals that resemble the remaining (defended) group of prey (through the emission of equally conspicuous visual signals) but who beyond the level of the strategy they play exhibit no inherent differences (such as in terms of their physiological adaptation to the defence) when compared with the defended type. While it may seem that this scenario is a simpler one to consider from the mathematical modelling perspective (only a single form for the fecundity F is considered) it also necessitates the use of a functional form for K that is not monotone - see (5.3.1). Related to this is the perhaps restrictive assumption that the model aversiveness is selected precisely at the

maximum value of K, which occurs at $t_1 = 1/\sqrt{k}$. The described scenario is not inaccurate and indeed it is interesting to observe that a prey population can exhibit an eco-evolutionarily stable dimorphism on the level of the defences - in Darst et al. (2006) it explained that once aposematism has become established in a population of prey stable levels of signalling and defence may become dissociated to allow for this. Nonetheless, the non-accounting of inherent inter-type differences is an omission that perhaps renders the setup less suitable for modelling mimicry, since mimicking types (Batesian or automimics) are expected to be different on some level, especially if these belong to a different species altogether. In keeping with a presentation that is progressively more elaborate (this also extends to the next chapter) we account for such differences on the level of the fecundity in the section that follows.

5.4 A stable continuum of solutions

In this section we explore an example in which the model and the mimic are assumed to vary not only in terms of their sensitivity to investment in secondary defences but also in terms of their base-line rates of reproduction. The functional forms that we consider are as follows

$$F_1(t) := f_{01}\exp(-f_1 t); \quad F_2(t) := f_{02}\exp(-f_2 t) \quad H(t) := t - t_c; \quad K(t) := \frac{k_0}{1+kt};$$

$$L(r) = D(r); \quad Q(i) := q_0\exp(-q\sqrt{I}); \quad S(x) = max(1-vx, 0). \tag{5.4.1}$$

We set $c := f_{01}/f_{02}$ to describe the base-line rates of fecundity as a proportion and remark that F_1 and F_2 represent the fecundities of the model and the mimic respectively. In the analysis we typically assume that while the mimic has a higher base-line for fecundity such that $f_{01} < f_{02}$ that this comes at a price since it is assumed that investment in defences is costlier to the mimic than it is to the model so that $f_1 < f_2$. These restrictions do not reflect hard rules but rather are grounded on the idea that the defended type is better adapted to secondary defences compared with its mimicking counterpart.

Conditions for eco-evolutionary stability

The aversive model (with $t_1 > t_c$) must satisfy equality (5.2.6) alongside inequality (5.2.7), which considering (5.4.1) amount to

$$-f_1 + \frac{k}{1+kt_1} + aq\frac{N}{n}D(r)\times\frac{1}{2\sqrt{I_1'}} = 0. \tag{5.4.2}$$

and

$$-f_1^2 + 2\left(\frac{k}{1+kt_1}\right)^2 + \frac{2k(f_1-k+f_1kt_1)}{(1+kt_1)^2} + \frac{(f_1-k+f_1kt_1)^2}{(1+kt_1)^2}$$

$$\frac{2k(f_1-k+f_1kt_1)^2\left[kt_1-(1+kt_1)\log\left(c+ckt_1\right)\right]}{(1+kt_1)^3(f_1t_1-\log\left(c+ckt_1\right))} > 0. \tag{5.4.3}$$

Much like in the previous section we consider a totally undefended mimic with $t_2 = 0$ so that inequality (5.2.8) for the mimic reads

$$-f_2 + k + aq\frac{N}{n}D(r)\times\frac{1}{2\sqrt{I_2'}} < 0. \tag{5.4.4}$$

The condition for ecological equilibrium (5.2.3) necessitates that the (resident) payoffs to the model and mimic are equal so that form (5.4.1) we have

$$\sqrt{I_2'} = \sqrt{I_1'} - \frac{f_1 t_1 - \log(c + ckt_1)}{q}. \tag{5.4.5}$$

The aversive information of the model and the mimic are precisely as in (5.2.1) and (5.2.2) from which it is implied that

$$I_2' = I_1' - a\frac{N}{n}D(r_1)t_1 \tag{5.4.6}$$

and indeed that

$$I_1' = \frac{N}{n}D(r)\big[t_1 - t_c - (1-a)\gamma t_1\big]. \tag{5.4.7}$$

We have argued in generality that a mimicry complex that is perceived collectively as attractive is unstable and in addition we impose that the perceived aversiveness of the mimics exceeds a certain lower bound I_ε of aversiveness, such that

$$\frac{N}{n}D(r_1)\left[(1-a)(1-\gamma)t_1 - t_c\right] > I_\varepsilon, \tag{5.4.8}$$

where a precise value of I_ε will be considered in the discussion that follows.

With functions (5.4.1) conditions (5.2.9) and (5.2.10) for invasion along r of the model amount to

$$-2\sqrt{I_1'}\frac{D'(r_1)}{D(r_1)} + aq\frac{N}{n}D'(r_1)(t_1 - t_c) + v(1-a)q\frac{N}{n}D(r_1)\left[(1-\gamma)t_1 - t_c\right] > 0 \tag{5.4.9}$$

and

$$-2\sqrt{I_1'}\frac{D'(r_1)}{D(r_1)} + aq\frac{N}{n}D'(r_1)(t_1 - t_c) - v(1-a)q\frac{N}{n}D(r_1)\left[(1-\gamma)t_1 - t_c\right] < 0. \tag{5.4.10}$$

For the mimic conditions (5.2.11) and (5.2.12) these are

$$-2\sqrt{I_2'}\frac{D'(r_1)}{D(r_1)} - aq\frac{N}{n}D'(r_1)t_c + v(1-a)q\frac{N}{n}D(r_1)\left[(1-\gamma)t_1 - t_c\right] > 0, \tag{5.4.11}$$

and

$$-2\sqrt{I_2'}\frac{D'(r_1)}{D(r_1)} - aq\frac{N}{n}D'(r_1)t_c - v(1-a)q\frac{N}{n}D(r_1)\left[(1-\gamma)t_1 - t_c\right] < 0. \tag{5.4.12}$$

Now that the conditions necessary to achieve eco-evolutionary stability in the mimicry complex have been identified we proceed to reducing these into a set of restrictions that can be readily manipulated numerically.

A strategy for determining eco-evolutionarily stable outcomes

Re-arrangement of (5.4.2) reads

$$\sqrt{I_1'} = \frac{aq\frac{N}{n}D(r_1)(1+kt_1)}{2\big(f_1 - k + f_1 kt_1\big)} \tag{5.4.13}$$

Squaring both sides of (5.4.5), using the latter re-arrangement of (5.4.2) and equality (5.4.6) leads to the following equality

$$a\frac{N}{n}D(r_1)\left[\frac{(1+kt_1)(f_1t_1 - \log(c+ckt_1))}{f_1 - k + f_1kt_1} - t_1\right] = \frac{(f_1t_1 - \log(c+ckt_1))^2}{q^2}, \tag{5.4.14}$$

which can be solved for $D(r_1)$ to give

$$D(r_1) = \frac{\big(f_1 t_1 - \log(c + ckt_1)\big)^2 (f_1 - k + f_1 kt_1)}{aq^2 \frac{N}{n}\big[kt_1(1 - \log(c + ckt_1)) - \log(c + ckt_1)\big]}. \tag{5.4.15}$$

Observing that the RHS of the latter depends only on the resident level of defence suggests that for some value of this is a unique associated level of the conspicuousness, which is given as the D pre-image of the vale on the RHS of (5.4.15).

We can work towards uncovering the optimal value of γ by first re-arranging (5.4.7) to give

$$\gamma = \frac{1}{1-a} \times \frac{t_1 - t_c}{t_1} - \frac{1}{1-a} \times \frac{I_1'}{\frac{N}{n} D(r_1) t_1}. \tag{5.4.16}$$

We can specify a unique value for the optimal size of the mutant group, by re-arranging (5.4.7) and substituting this into (5.4.15) such that

$$\gamma = \frac{1}{1-a} \times \frac{t_1 - t_c}{t_1} - \frac{a}{1-a} \times \frac{\big(f_1 t_1 - \log(c + ckt_1)\big)^2 (1 + kt_1)^2}{4(f_1 - k + f_1 kt_1)\big[kt_1(1 - \log(c + ckt_1)) - \log(c + ckt_1)\big] t_1}. \tag{5.4.17}$$

Having obtained $D(r_1)$ and γ as expressions involving the model defence t_1 only, we can express (5.4.8) as an inequality in the variable t_1 only. Indeed, substitution of (5.4.15) and (5.4.17) amounts to

$$\frac{(f_1 t_1 - \log(c + ckt_1))^2 (f_1 - k + f_1 kt_1)}{q^2 \left[kt_1 - (1 + kt_1)\log(c + ckt_1)\right]} \times \left[0.25 \frac{(f_1 t_1 - \log(c + ckt_1))^2 (1 + kt_1)^2}{(f_1 - k + f_1 kt_1)\left[kt_1 - (1 + kt_1)\log(c + ckt_1)\right]} - t_1\right] > I_\varepsilon.$$

It remains for us to explore the consequences of (5.4.15), which can be re-arranged to read

$$\sqrt{I_2'} > \frac{aq\frac{N}{n} D(r)}{2(f_2 - k)}, \tag{5.4.18}$$

where we notice that the LHS can be re-expressed through (5.4.5) as

$$\sqrt{I_1'} - \frac{f_1}{q} t_1 + \frac{\log(c + ckt_1)}{q} > \frac{1}{2} \times \frac{aq\frac{N}{n} D(r)}{f_2 - k}. \tag{5.4.19}$$

Substitution of (5.4.13) for $\sqrt{I_1'}$ results in the inequality

$$aq\frac{N}{n} D(r)\left[\frac{1 + kt_1}{f_1 - k + f_1 kt_1} - \frac{1}{f_2 - k}\right] - \frac{2f_1}{q} t_1 + \frac{2}{q}\log(c + ckt_1) > 0, \tag{5.4.20}$$

which through substitution of (5.4.15) gives rise to inequality

$$\frac{(f_1 t_1 - \log(c + ckt_1))^2\big[kt_1(f_2 - f_1 - k) + f_2 - f_1\big]}{(f_2 - k)(kt_1 - (1 + kt_1)\log(c + ckt_1))} - 2f_1 t_1 + 2\log(c + ckt_1) > 0, \tag{5.4.21}$$

which is equivalent. From (5.4.5) and (5.4.6) it follows that $f_1 t_1 - \log(c + ckt_1) > 0$ and by factoring out this term we deduce that the latter inequality is equivalent to

$$[f_1 t_1 - \log(c + ckt_1)]\,[kt_1(f_2 - f_1 - k) + f_2 - f_1] - 2(f_2 - k)(kt_1 - (1 + kt_1)\log(c + ckt_1)) > 0. \tag{5.4.22}$$

Through suitable re-arrangement this inequality amounts to

$$f_1k(f_2-f_1-k)t_1^2+k(f_1+f_2-k)\log(c+ckt_1)t_1+(f_1f_2-f_2-2kf_2+2k^2)t_1>0, \tag{5.4.23}$$

which we make use of in the discussion that follows.

As far as solving the conditions (5.4.9), (5.4.10) and (5.4.11), (5.4.12) we observe the following: the factor $(1-\gamma)t_1-t_c$ in the third terms is positive then either pair of inequalities ((5.4.9) - (5.4.10) and (5.4.11) - (5.4.12)) can be satisfied providing the associated value for v is chosen to be large enough. Since we are working under the regime that $I_2'>I_\varepsilon$ (i.e. that both model and mimic are perceived by the predator as types that are aversive) this implies that

$$(1-\gamma)t_1-t_c>(1-a)(1-\gamma)t_1-t_c>0. \tag{5.4.24}$$

In particular, the first pair of inequalities (5.4.9),(5.4.10) describing the stability of the model along r can be satisfied providing a large enough v is chosen such that

$$v>\frac{\left|aq\frac{N}{n}D'(r)(t_1-t_c)-2\sqrt{I_1'}\frac{D'(r)}{D(r)}\right|}{(1-a)q\frac{N}{n}D(r)\big[(1-\gamma)t_1-t_c\big]}=:v_1 \tag{5.4.25}$$

and in a similar fashion the mimic inequalities (5.4.11) and (5.4.11) are satisfied providing

$$v>\frac{aq\frac{N}{n}D'(r)t_c+2\sqrt{I_2'}\frac{D'(r)}{D(r)}}{(1-a)q\frac{N}{n}D(r)\big[(1-\gamma)t_1-t_c\big]}=:v_2. \tag{5.4.26}$$

It therefore follows that if both model and mimic are perceived as aversive then both pairs of conditions (5.4.25) and (5.4.26) can be satisfied providing that

$$v>max\big(v_1,v_2\big). \tag{5.4.27}$$

Numerical analysis

At this stage we are ready to explore specific numerical examples. We begin by specifying the rate of detection D as

$$D(r_1)=\frac{0.01}{0.011+0.99\exp(-r_1)}\quad\text{with}\quad d_0=0.01. \tag{5.4.28}$$

In principle the ratio f_{01}/f_{02} can assume any positive value but we will assign to this the value $f_{01}/f_{02}=0.5$, suggesting that the base-rate of reproduction of the mimic is higher (specifically, twice the value) than that of the model. Equality (5.4.2) for the model requires that parameters f_1 and k must be chosen such that

$$f_1>\frac{k}{1+kt_1}$$

and in fact, we use the stronger bound $f_1>k$ and hence assign the values $f_1=3$ and $k=2$ to the parameters. Considering (5.4.15) it is clear that specifying the product aq^2N/n fully specifies the curve in (5.4.2) for the model. We set $aq^2N/n=80$ and obtain the following implicit equality, which for $r_1^*>0$ identifies level of

defence $t_1^*(r_1^*)$ uniquely as function that increases with respect to r_1^* such that

$$\frac{0.01}{0.01+0.99\exp(-r_1)}=\frac{\left(3t_1-\log\left(0.5+t_1\right)\right)^2\left(1+6t_1\right)}{80\left(2t_1-\left(1+2t_1\right)\log\left(0.5+t_1\right)\right)}. \tag{5.4.29}$$

A section of the curve in (5.4.29) is plotted in Figure 5.2. An initial lower and upper bound for t_1 are provided through (5.4.29) corresponding to $r_1=0$ and $r_1\to\infty$, since the LHS is bounded between $[0.01,1)$. Within this interval the LHS and RHS are individually monotonically increasing functions and therefore we can express the value t_1 not only implicitly through $r_1(t_1)$ but also explicitly as $t_1(r_1)$. The bounds are therefore defined through

$$t_1(r_1=0)\approx 0.0214 \text{ and } t_1(r_1)\approx 1.0765 \text{ for } r_1\gg 1. \tag{5.4.30}$$

As we establish, the latter is one of several competing lower-bounds on t_1, from which we ultimately consider the largest. As it happens, the upper bound in (5.4.30) will prove to be the least of these.

For a model strategy $(r(t_1),t_1)$ that solves (5.4.29) we must now identify a value of a and a value of $t_c<t_1$ such that there exists at least one associated value of γ within the interval $(0,1)$. Setting $a=0.2$ and $t_c=0.2$ we observe that (5.4.17) amounts to

$$\gamma(t_1)=1.25-\frac{0.25}{t_1}-0.125\frac{(3t_1-\log(0.5+t_1))^2(1+2t_1)^2}{(1+6t_1)\left[2t_1-(1+2t_1)\log(0.5+t_1)\right]t_1} \tag{5.4.31}$$

where for the range of values of $t_1\in(0.0214,1.0765)$ satisfying (5.4.29) it is true that

$$\gamma(t_1)>0 \;\Leftrightarrow\; t_1\in(0.2524,1.1599)\,. \tag{5.4.32}$$

The lower bound of (5.4.32) is larger than that of (5.4.30) and therefore imposes a stronger restriction on the lowest admissible values for the level of t_1. Meanwhile, the upper bound on t_1 is higher than that of (5.4.30) and can be disregarded.

It should be noted that for our choices of parameters f_1, k and c strategies (r_1,t_1) satisfying (5.4.29) also satisfy (5.4.3), which from (5.4.1) reads

$$-9+2\left(\frac{2}{1+2t_1}\right)^2+\frac{4\left(1+6t_1\right)}{\left(1+2t_1\right)^2}+\frac{\left(1+6t_1\right)^2}{\left(1+2t_1\right)^2}+\frac{4\left(1+6t_1\right)^2\left(2t_1-\log\left(0.5+t_1\right)\left(1+2t_1\right)\right)}{\left(1+2t_1\right)^3\left(2t_1-\log\left(0.5+t_1\right)\right)}>0. \tag{5.4.33}$$

As it happens this inequality is trivially true for all values of $t_1=t_1$ that satisfy (5.4.29). Implicit in this discussion has been the validity of the regime $I_2'>I_\varepsilon$ for some choice of I_ε. Setting $I_\varepsilon=0.05$ and $q=2$ (the latter also implies that $N/n=100$) we deduce that

$$\frac{\left(3t_1-\log\left(0.5+t_1\right)\right)^2\left(1+6t_1\right)}{4\left(2t_1-\left(1+2t_1\right)\log\left(0.5+t_1\right)\right)}\times\left(\frac{0.25\left(3t_1-\log\left(0.5+t_1\right)\right)^2\left(1+2t_1\right)^2}{\left(1+6t_1\right)\left(2t_1-\left(1+2t_1\right)\log\left(0.5+t_1\right)\right)}-t_1\right)>0.05, \tag{5.4.34}$$

which holds true for

$$t_1>0.4456.$$

The lower bound for t_1 of (5.4.34) is the largest encountered thus far and represents the smallest value of t_1 (although see discussion that follows) for which the curve assumes the form of (5.4.15). In fact, for values of

$t_1 < 0.4456$ the curve in (5.4.15) would have to be evaluated using a different branch of the Q-function. The scenario that is not explored presently is that in which $I_2' < I_\varepsilon < I_1'$. Should solutions be present in that regime, they should simply be considered jointly with those that we detail here and not influence them.

In order to address condition (5.4.23) for the mimic we must select a value for the parameter f_2 such that this is possible. Viewing (5.4.2) and (5.4.4) in tandem we infer that $f_2 > f_1$ and that $f_2 > k$. In this instance we have the ordering $f_2 > f_1 > k$. Indeed, this reflects that fact that although mimics have higher base-fecundity rate, their fecundity is more sensitive to marginal increases in their fecundity (the F_2 exhibits a steeper drop-off than the F_1 curve, even though the base-rate fecundity of the mimic is taken to be exactly double that of the mimic - see $c = 0.5$). Taking $f_2 = 5$ inequality (5.4.23) reads

$$12 \log (0.5 + t_1) - 2t_1 > 0 \;\; \Rightarrow \;\; t_1 > 0.612. \tag{5.4.35}$$

This is the greatest lower bound on t_1 yet and is represented as the top dash-dotted vertical line in the plots of Figure 5.2. This revised range of admissible defence given approximately as $t_1 \in (0.612, 1.0765)$ also specifies the range of admissible conspicuousnesses through (5.4.15). In particular, the lower bound $t_1 \approx 0.612$ corresponds to the lower bound of conspicuousness given approximately as $r_1 \approx 3.0577$, while the upper bound of the defence $t_1 \approx 1.0765$ is achieved asymptotically as $r_1 \to \infty$. As is clear from Figure 5.2(b) the equilibrium mimic proportion γ provided through (5.4.17) has a local maximum of $\gamma \approx 0.5179$ at $t_1 \approx 0.6586$ and decreases monotonically for larger values of t_1 so that γ assumes values within $(0.188, 0.5179]$. We establish that eco-evolutionarily stable solutions for $(t_1, r_1(t_1), \gamma^*(t_1))$ should exist within the 3-rectangle $(0.612, 1.0765) \times (3.0577, +\infty) \times (0, 188, 0.5179]$ of the extended strategy space.

We have demonstrated that model strategies (r_1, t_1) drawn from the curve (5.4.29) are stable against invasion along t and for each strategy on the continuum exists an associated mimic that is undefended and in turn also can resist invasion along t on account of (5.4.35). The mimic resides in proportion γ as given in (5.4.31) and for that level shares the same payoff as the mimic such that the payoff balance is stable in small perturbations around the mimetic load in the sense of (5.1.17).

It remains for us to determine whether there is a subset of the strategy space on which the model and mimic are resist invasion along r and in particular whether there exists a choice for parameter v such that the two pairs of inequalities (5.4.9) through to (5.4.12) can be satisfied over some non-empty subset of the strategy space. Experience tells us that these inequalities can be solved for a value of v that is sufficiently large - reflecting the intuition that a careful enough predator will punish visually dissimilar mutants sufficiently to make the type-1 and type-2 residents safe against potential invaders of this type. Indeed, the assumption that both model and mimic are perceived as aversive allows us to solve (5.4.9), (5.4.10) and (5.4.11) and (5.4.12) as pairs of inequalities by choosing a v that is large enough - see (5.4.25) through to (5.4.27). While choosing a very large value for v is sure to achieve this result we seek plausible values for this parameter by finding sensible bounds - see v_1 and v_2 below.

We observe that v_1 and v_2 in (5.4.25) have a common denominator, which we can bound from below

$$\begin{aligned}(1-a)q\frac{N}{n}D(r_1)[(1-\gamma)t_1 - t_c] &> 0.8 \times 2 \times 100 \times D(r_1 \approx 3.0576) \times [(1 - \gamma(t_1 \approx 0.6586)) \times 0.612 - 0.2] \\ &\approx 2.690.\end{aligned}$$

The numerator in the expression for v_2 provided in (5.4.26) can be bounded from above as follows

$$\begin{aligned} aq\frac{N}{n}D'(r_1)t_c + 2\sqrt{I_2'}\frac{D'(r_1)}{D(r_1)} &< 0.2 \times 2 \times 100 \times 0.2 \times D'(r_1 \approx 3.0576) \\ &\quad + 2 \times \sqrt{I_2'(t_1 \approx 1.0765)} \times \frac{D'}{D}(r_1 \approx 3.0576) \\ &\approx 1.165 + 11.632 \approx 12.797, \end{aligned} \tag{5.4.36}$$

so that

$$v_2 < \frac{12.797}{2.690} \approx 4.757. \tag{5.4.37}$$

We use a similar process of bounding for v_1; we keep the lower bound on the denominator of 2.690, since this is common to both expressions (5.4.25) and (5.4.26) and proceed to finding an upper bound on the numerator. Noticing that the numerator involves the absolute difference of two terms, it is not immediately clear which of these has a leading maximum. Making use of (5.4.5) we deduce that the first term $2\sqrt{I_1'}D'(r_1)/D(r_1)$ is bounded between

$$0 \approx 2 \times \sqrt{I_1'(t_1 \approx 0.612)} \times \frac{D'}{D}(r_1 \to \infty) < 2\sqrt{I_1'}\frac{D'(r_1)}{D(r_1)} < 2 \times \sqrt{I_1'(t_1 \approx 1.0765)} \times \frac{D'}{D}(r_1 \approx 3.0576) \approx 2.594$$

and that for the second term it is true that

$$0 \approx 40 \times D'(r_1 \to \infty)(0.612 - 0.2) < aq\frac{N}{n}D'(r_1)(t_1 - t_c) < 40 \times D'(r_1 \approx 3.0576) \times (1.0765 - 0.2) \approx 5.105.$$

Therefore,

$$v_1 < \frac{5.105}{2.690} \approx 1.898. \tag{5.4.38}$$

In particular, choosing $v = 5$ guarantees that the solutions of Figure 5.2 satisfy inequalities (5.4.9) through to (5.4.12). The results of the numerical analysis can be seen in Figure 5.2 in which is shown that models and mimics can co-exist in a sense that is stable both from the (long-term) evolutionary time-scale and from the short-term ecological time-scale providing the model aversiveness is drawn from the black section of the curve in Figure 5.2(a) and the associated mimetic load is drawn from the black section of the solid curve in 5.2(b). Figure 5.2(a) shows that stable levels of model defence are observed to increase with increasing levels of conspicuousness, although a plateau is observed such that beyond a certain level the associated stable level of defence does not change. Interestingly, in Figure 5.2(b) we observe that while initial increases in conspicuousness are associated with higher levels of stable mimetic loads the relationship is reversed beyond a certain point so that the more conspicuous complexes consist increasingly of models.

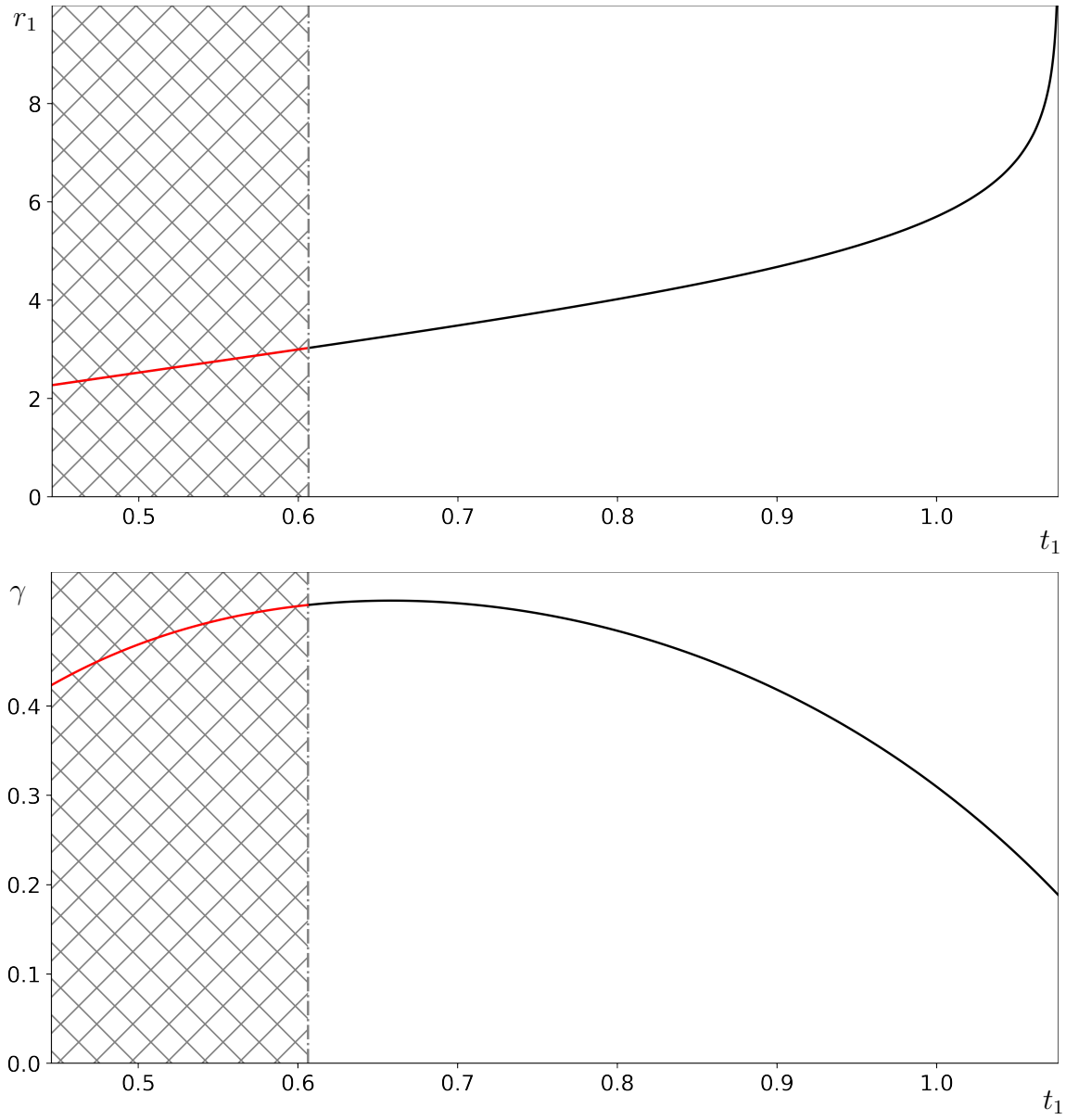

Figure 5.2: The shaded regions span the interval $0.4456 < t_1 < 0.612$ in condition (5.4.35) for the mimic fails - the curves mentioned in (a) and (b) are drawn in red over these regions. The region $t_1 < 0.4456$ is not shown because on this $I_2' < I_\varepsilon$, which is also a violation. **(a)** [Top] Solid curve shows conspicuousness versus model defence as provided implicitly through (5.4.29); the red section is unstable in the sense of failing (5.4.35) while the black section beyond $t_1 \approx 0.612$ admits a continuum that is eco-evolutionarily stable when viewed in tandem with the continuum in (b). **(b)** [Bottom] Solid curve shows the mimetic load vs. model defence as provided explicitly through (5.4.31); as mentioned earlier only the black section on $t_1 > 0.612$ admits an equilibrium that is eco-evolutionarily stable. Plots (a) and (b) can be viewed in tandem and thus perceived as a single curve $(t_1, r(r_1), \gamma(t_1))$ defined through (5.4.29) and (5.4.31) within the subregion $(0.612, 1.0765) \times (3.0577, +\infty) \times (0, 188, 0.5179]$ of the extended strategy space.

An alternative approach to modelling mimicry

In the previous section we demonstrated that the mimetic co-existence of two types that differ only on the level of the strategies they play and the proportions in which they do so can be achieved and hence argued that the specified solution likely lies on a continuum, which we did not explore. This example served more as

an introduction to the study of mimicry within the framework of evolutionary stability and perhaps models a situation that is rather idealised; in most instances of mimicry be it Batesian or automimicry inherent differences between the model and the mimic are likely present. Furthermore, while in our first approach we made use of fewer functional forms (by setting $F_1 = F_2$) we saw that more careful consideration of the form for K had to be taken to make the system stable. In this section we accounted for type-specific differences by considering two sets of functional forms for F so that the types are assumed to differ with respect to how a given level of defence impacts their fecundity. In the latter, we were able to demonstrate that eco-evolutionarily stable solutions are manifest as an increasing continuum in the (extended) strategy space (t_1, r_1, γ) that includes the conspicuousness, model defence and proportion of mimics. While these results certainly appear promising there are a number of areas that could have been addressed differently to model mimicry complexes involving two distinct species.

In this chapter we have assumed that there is some background proportion of mimics, which is given by γ such that in every site one is expected to encounter approximately $(1-\gamma)N$ models and γN mimics. Much like in the single-species description of chapters 2,3 and 4 mutation is facilitated by means of local clustering so that in a small number of sites there are nests of size a containing relatives of a certain focal relative, which could play a strategy that is local to the type in question. The definitions of the mutant perceived aversiveness as introduced in (5.1.2) suggests that the parameter a measures the concentration of focal relatives as a proportion over the total number of prey residing in that site, which in turn suggests that the proportion of models and mimics is perturbed in the given site. It can be argued that if the types were of a different species that such an assumption would be less plausible. In the chapter that follows we model the proportion of mimics as a continuous (beta-distributed) random variable, which is perhaps a more realistic assumption in the sense that while it is true that on average one expects to encounter a certain mean proportion of mimics in a given site, there is considerable variation around this value. In addition, while mutation is still facilitated through local clustering parameter a measures the proportion of mutants with respect to the number of individuals of that type so that the presence of type-1 mutants does not influence the proportion of type-2s in that site and likewise the presence of type-2 mutants does not influence the proportion of type-1s. In this way the presence of, say type-1 mutants impacts the proportion of type-1 residents to type-1 mutants without affecting the proportion of type-1s to type-2s.

One could argue that our understanding of the local relatedness parameter as provided in this chapter best describes a scenario in which the two types belong to the same species, while the description provided in the next chapter better describes types belonging to different species. The approach of the next chapter is more general and while it may best describe Batesian mimicry it may be applicable to a broader range of mimicry complexes; on the other hand it is likelier that the results in this chapter are better suited to model automimicry systems but perhaps not to the modelling of Batesian mimicry systems. Furthermore, while we have accounted for inherent differences between the types of prey in this closing section we have done so only from the point of view of the fecundity and neglected that the impact of investment in defence on the escaping attacks is also a type-specific property; in the chapter that follows we account for this too by considering two sets of forms for K, the probability that attack results in capture. While the approach of this chapter consists of a more straightforward adaptation of the single-species description of aposematism it is also easier to work with and perhaps more straightforward to validate or not a certain conclusion with respect to the underlying assumptions; we discover in the chapter that follows that the numerical analysis is notably more involved. We also leave discussions of the literature to the end of the chapter that follows for the reason that these can more confidently be related to observed instances of mimicry.

Chapter 6

Batesian Mimicry

Mimicry systems are manifest in a large number settings in the natural world and common to a broad range of taxa. Even though, in practice, the various kinds may not be straightforward to distinguish, we clarify that the content of this chapter is specifically dedicated to *deceptive* and *protective* forms of Batesian mimicry (although the reader may find its relevance to *Browerian* mimicry, its automimicry analogue). We discuss some biological background, including explanations of some key details in the first section but encourage the reader to consult chapters 9 of Ruxton et al., (2019) for a more complete and organised overview.

The mathematical modelling of Batesian mimicry is considered from the second section and onwards. It commences with a bridging of the model by Broom et al. (2006) from the single to the two-species case; initially the presentation draws on ideas coming from sampling theory and is quite general in the sense of it (perhaps) being more relevant to real, finite populations that can be studied numerically. In the third section the same setting is re-introduced and expanded, but now with with the adoption of more abstract notions from probability theory (such as the formal assumption that the proportion of mimics-to-models is a randomly distributed variable). The fourth section deals with ecological stability, while in the fifth we expand our usual conception of a local ESSs to describe the co-evolutionary stability of the model and mimic as a complex. In the sixth section we consider specific functional forms for the predator generalisation and express the conditions for bi-stability (ecological and evolutionary) in terms of the confluent hypergeometric function. In the closing section we consider specific functional forms for the prey and provide a working example of a Batesian mimicry system that is eco-evolutionarily stable in the sense discussed. We contend that our findings, especially of this last section are novel; we discuss their relevance with existing bodies of work and address areas for future consideration.

6.1 Background into Batesian mimicry

In this section we provide some essential non-mathematical background relating to Batesian mimicry. It is beyond our scope to provide a through description of the process here - the reader is directed to chapter 9 of Ruxton et al. (2019) and the vast bibliography sited therein for this. We focus our attention to four themes, which we find are most relevant to the mathematical analysis that follows. Namely (i) what are the distinguishing characteristics of a Batesian mimicry system and how do they "work"? (ii) Does the success of a mimic necessitate the presence of a model? If so, how do (iii) the noxiousness of the model and (iv) the proportion of mimics to models influence the success of the mimic?

Mimicry (or deceptive mimicry) can be broadly understood as a means of deception for the gaining of a selective advantage. There are a number of forms in which mimicry systems can arise and these can be challenging to classify in practice. Nonetheless, there is now ample evidence to suggest that individuals from one species may gain selective advantage by mimicking the appearance of individuals from another species, with the example of Kirby and Spence (1817) now dating over 200 years. The first to conceive of mimetic resemblance as a potential mechanism for selection was Henry Walter Bates in his prolific work - Bates (1862) - on Amazonian butterflies. Although that study examined various different forms of mimicry he is better known for the form we now identify as *Batesian* mimicry, in which individuals from a palatable species (*mimics*) gain protection from predators by resembling an unpalatable (or otherwise defended species) known as the *models*.

Batesian mimicry is a form of *protective* mimicry as it is thought to have evolved to provide prey with protection against predators. It is to be contrasted with *aggressive* mimicry in which individuals from one species resemble individuals from another to gain access to prey, hosts, or other resources (see Ruxton et al, 2019 for more on this). Forms of aggressive mimicry are altogether not considered here. We mention three related forms of protective mimicry and underline their key differences. These are: (a) *Masquerade*, (b) *Müllerian* mimicry and (c) *Browerian* mimicry. As we explain (a) and (b) are not accounted for in our mathematical description, while (c) is.

Masquerade is the resemblance of a species to individuals or objects (such as leaves, twigs or stones) that are of no inherent interest to the predator (see Skelhorn, 2015 and/or Skelhorn et al., 2010). It may appear that masquerade is indistinguishable from crypsis, which has concerned an important portion of the manuscript. As explained in Allen and Cooper (1985) crypsis is an adaptation that reduces the chance of predator *detection*, while masquerade reduces the chance of predator *recognition* of prey (as a nutrient-rich items) without necessarily influencing their detectability. (b) *Müllerian* mimicry (named after Johannes Friedrich Müller - see Müller, 1879) describes the evolution of a shared warning signal among different unpalatable species. While Batesian mimics are palatable species that resemble less palatable ones, Müllerian mimics are unpalatable species that resemble other unpalatable ones to achieve a potentially greater protection against predators than they would through individual signals. It is speculated that this shared signal could itself serve as a stabilising mechanism by better facilitating aversive learning in predators.

Finally, (c) Browerian mimicry (named after Lincoln Pierson and Ane Van Zandt Brower) can be thought of as the *automimicry* (coined in the classic study by Brower et al., 1967) equivalent of Batesian mimicry in the sense that palatable mimics resemble unpalatable individuals that belong to the same species. Browerian mimicry is present among populations of a given species that exhibit a spectrum of palatability - see Gamberale-Stille and Guilford (2004); Skelhorn and Rowe (2007); Jones et al. (2013) for examples on bird-insect systems including chemically-defended prey and the review by Speed et al., (2012) discussing automimicry in a number of different taxa - and are in that respect intrinsically different to Batesian mimicry complexes. As discussed in Brower et al., (1967) for instance, the noxiousness of individual monarch butterflies (*Danaus plexippus*) depends on the availability of "*cardenolide secondary metabolites*" that are acquired from the host plant during the larval stage of development. Therefore, while in some instances automimicry may result from a natural variability in a specie's palatability the evolution of Batesian mimicry may involve a number of more complex mechanisms. Interestingly, we treat much of their mathematical modelling in an all-inclusive way such that the distinction between Batesian mimics and automimics only becomes relevant when it comes to choosing exact prey functional forms (see section 6.6 some 20 pages following forward).

So far we have provided some descriptive background relating to Batesian mimicry (what it is, what it

is not, how it related to other forms etc.). One obvious question, which we address presently is, *does it work?* As mentioned earlier, Batesian mimicry relies on the successful deception of predators and indeed there is ample empirical evidence to suggest that this does happen. Laboratory experiments (such as Mostler (1935), or the three-volume studies Brower, 1958,a,b,c or Platt et al., 1971 or the more recent Kuchta et al., 2008) provide direct evidence of this. Possibly due to the larger number of inherent complexities associated with experiments conducted in the wild, laboratory studies are more conclusive with respect to this *duping* effect compared with associated field studies. Early field studies, such as those by Brower et al. (1964), Brower et al. (1967) and Cook et al. (1969) are remarkably fascinating but have arguably reported mixed results. The study by Jeffords et al. (1979) was (perhaps accurately) described by the authors as the first field experiment to verify the efficacy of Batesian mimicry, which when viewed in tandem with Waldbauer and LaBerge (1985) could suggest that the success of the mimic could necessitate the presence of a noxious model. We elaborate on the latter presently.

Both Bates and Wallace were strong advocates of the hypothesis that the role of the presence of the model for the success of the Batesian mimic is vital. In Pfennig et al. (2001) and later also in Pfennig et al. (2007) it was demonstrated that attacks on (artificial) mimics were higher in locations without models compared with those in which models were also present. But there is also intriguing indirect evidence of the importance of the presence of the model, which can be made out through the forms mimics assume in those areas in which models are and those in which models are not present (see the works of Hecht and Marien, 1956 and Edmunds and Edmunds, 1974 on the non-venomous *Lampropeltis doliata* snake, which mimics the *Erythrolamprus aesculapii* snake). M. Edmunds and Reader (2014) observed that the relative frequency of the black and yellow hoverfly mimic (*Volucella bombylans plumata*) was positively related to the relative frequency of the black and yellow bumblebee model. Overall it seems important to the maintenance of mimicry for model and mimic to co-occur in a certain geographical location, although interestingly these may not have to co-occur at the same time (see Rothschild, 1963, Brodie, 1981 or Waldbauer and LaBerge, 1985 for seasonal prey). Very exceptional cases of mimics that exist without models do exist (see Clarke and Sheppard, 1975), but these remain the overwhelming exception.

As we have already established in earlier chapters an important mechanism for the maintenance of aposematism (and perhaps of mimicry) is aversive learning, which broadly describes the formation of cognitive associations from the point of view of the predator with respect to encountered prey. It would be natural to presume that, *ceteris paribus*, the success of the mimic is increasingly likely as it resembles increasingly noxious models. Laboratory experiments by Skelhorn and Rowe (2006) showed that domesticated chick predators are reluctant to consume palatable food items after having been previously subjected to identical foods injected with quinine sulphate. This intuitive prediction has also been confirmed by similar laboratory studies such as that by Alcock (1970), which showed that white-throated sparrows (*Zonotrichia albicollis*) attacked neither (mealworm) model or mimic when the latter was sufficiently aversive. The study by Lindström et al. (1997) showed that great tit predators attacked mealworm prey less when the model chloroquine concentrations were highest although these were able to distinguish mimics that were imperfect and attack them at the same rate.

This brings us onto the fourth and final theme of this introductory section, namely *in what way does the success of the Batesian mimic depend on its relative abundance in the complex?* An educated guess would tell us that there is a continuously decreasing relationship between the (model and mimic) fitnesses with respect to increasing mimic proportions in the population, perhaps due to a sequential reduction in the predator's perceived aversiveness of the complex. A study conducted by Brower (1960) on European

starling predators (*Sturnus vulgaris*) showed that as the proportion of model mealworms (immersed in quinine dihydrochloride) increased the rates of attack on both the model and mimic decreased. Notably this decrease was non-linear so that there was no observed increase in the rate of attack as the proportion of models decreased marginally below a certain threshold. The laboratory study by Jones et al. (2013) on bird predators feeding on mealworm larvae consisting of either edible mimics or unpalatable models confirmed both these findings. Namely, through varying the proportion of so-called *cheats* from 0 to 1 by increments of 0.25 they discovered that the proportion of prey attacked increased non-linearly with the frequency of cheats. A negative (and non-linear) relationship between Batesian mimic fitness and relative abundance was also established through a series of Monte Carlo simulations in Turner et al. (1984).

6.2 A sampling theory approach

We begin by considering a habitat that is occupied by two distinct types of prey and which is *territorially divided* among the predators who visit it. In particular, we consider a habitat partitioned into M distinct localities (indexed by $i = 1, 2, ..., M$), each consisting of precisely N prey. Within these there is a proportion $(1-\gamma_{[i]})$ of prey individuals who are aversive (type-1) and play strategy (r_1, t_1) with $t_1 > t_c$ and a proportion $\gamma_{[i]}$ of individuals (type-2) who reproduce the signal of the first but who do not invest in toxins at all; these individuals play strategy $(r_1, 0)$.

Territorial division of the habitat is in this context the assumption that the predator population is divided into distinct sets of size n who each visit a specific locality only (this is a direct extension of the single-species description considered thus far). Overall, it is presumed that each locality has the capacity to carry a large enough number of prey, such that the approximation $N \gg 1$ is thought to hold (see calculations below - although strictly speaking the limit $N \to \infty$ is considered at a separate step). It should also be remarked that the above description is one of a *perfect mimicry system* in that the *mimic* (type-2) reproduces the signals of the *model* (type-1) exactly.

The prey individuals within a locality are enumerated with index j and we may order these for convenience such that the index values $j = 1, ..., \text{nint}\left((1-\gamma_{[i]})N\right)$ correspond to type-1 individuals, while remaining indices $j = \text{nint}\left((1-\gamma_{[i]})N\right) + 1, ..., N$ are assigned to type-2's. Schematically, we write

$$j = \underbrace{\overbrace{1}^{\text{focal}}, 2, ..., \text{nint}\left((1-\gamma_{[i]})N\right)}_{\text{type-1}}, \underbrace{\overbrace{\text{nint}\left((1-\gamma_{[i]})N\right)+1}^{\text{focal}}, \text{nint}\left((1-\gamma_{[i]})N\right)+2..., N}_{\text{type-2}} \quad i = 1, ..., M. \quad (6.2.1)$$

We make use of the *nearest integer function* as defined in WolframAlpha (2022), such that nint(x) returns the nearest integer to $x \in \mathbb{R}$ with odd half-integers being rounded up and even half-integers are rounded down. This type of rounding function is preferable to the likes of $\lceil x \rfloor$ used in Hastad et al. (1989), as it uses the simpler rule $\lceil x \rfloor = x - 0.5$ when x is a half-integer, which could lead to statistical biasing. In the calculations that follow nint is omitted so as to avoid making the notation excessively tortuous; the reader is encouraged to think of $(1-\gamma_{[i]}N)$ as $\text{nint}\left((1-\gamma_{[i]}N)\right)$ etc..

Aversive information

We are interested in determining the aversiveness of a type-1 individual as perceived by the average predator visiting its site. We consider some focal individual from the type-1 subgroup (we pick $j = 1$ from the

ordering introduced in (6.2.1)) and apply definition (2.2.2) from the model description in section 2. In doing so, we implement a modified notation with subscripts in square brackets $\{[i,j] : i = 1, ..., M; j = 1, ..., N\}$ to indicate that the quantity in question is evaluated for individual j of site i. With the ordering of (6.2.1) in place, the aversiveness of $j = 1$ in site i reads

$$\begin{aligned} I_{[i,j=1]} &= \frac{L(r_1)}{n}\sum_{j=2}^{N} H(t_j) \\ &= \frac{L(r_1)}{n}\sum_{j=2}^{(1-\gamma_{[i]})N} H(t_j) + \frac{L(r_1)}{n}\sum_{j=(1-\gamma_{[i]})N+1}^{N} H(t_j) \\ &= \frac{(1-\gamma_{[i]})N-1}{n} L(r_1)H(t_1) + \gamma_{[i]}\frac{N}{n}L(r_1)H(0). \end{aligned} \tag{6.2.2}$$

We work in a similar fashion to determine the aversiveness of a type-2 individual as perceived by the average predator visiting its site. That is, we select focal individual $j = (1-\gamma_{[i]})N + 1$ from site i using the enumeration in (6.2.1) to write

$$\begin{aligned} I_{[i,j=(1-\gamma_{[i]})N+1]} &= \frac{L(r_1)}{n}\sum_{\substack{j=1 \\ j\neq(1-\gamma_{[i]})+1}}^{N} H(t_j) \\ &= \frac{L(r_1)}{n}\sum_{j=1}^{(1-\gamma_{[i]})N} H(t_j) + \frac{L(r_1)}{n}\sum_{j=(1-\gamma_{[i]})N+2}^{N} H(t_j) \\ &= (1-\gamma_{[i]})\frac{N}{n}L(r_1)H(t_1) + \frac{\gamma_{[i]}N-1}{n}L(r_1)H(0). \end{aligned} \tag{6.2.3}$$

It should be clear that the RHS of (6.2.2) does not depend on the choice of focal individual from within the type-1 subgroup, such that

$$I_{[i,j]} = \frac{(1-\gamma_{[i]})N-1}{n}L(r_1)H(t_1) + \gamma_{[i]}\frac{N}{n}L(r_1)H(0) \text{ holds for all } j = 1, ..., (1-\gamma_{[i]})N \tag{6.2.4}$$

holds. The same reasoning applies for the RHS of (6.2.3) and so we write

$$I_{[i,j]} = (1-\gamma_{[i]})\frac{N}{n}L(r_1)H(t_1) + \frac{\gamma_{[i]}N-1}{n}L(r_1)H(0) \text{ holds for all } j = (1-\gamma_{[i]})N+1, ..., N. \tag{6.2.5}$$

Equalities (6.2.4) and (6.2.5) suggest that our subscript notation $[i,j] : i = 1, ..., M; j = 1, ..., N$ is superfluous. Instead, we modify the subscript notation so that the second entry denotes only the type number of the individual in site i, i.e. $\{[i,\pi]; i = 1, ..., M; \pi = 1, 2\}$. With this simpler notation in mind we write the

aversiveness for the average type-1 individual in site i as

$$I_{[i,1]} := \frac{(1-\gamma_{[i]})N-1}{n} L(r_1)H(t_1) + \gamma_{[i]}\frac{N}{n}L(r_1)H(0) \tag{6.2.6}$$

and for the average type-2 individual in site i as

$$I_{[i,2]} := (1-\gamma_{[i]})\frac{N}{n}L(r_1)H(t_1) + \frac{\gamma_{[i]}N-1}{n}L(r_1)H(0). \tag{6.2.7}$$

Simulations of the model for finite populations may call for expressions (6.2.6) and (6.2.7) in which exclusion of the focal individual from the remaining population in the site is important.

We keep such ideas open and remark that for a large enough number of prey occupying a given site (i.e. $N \gg 1$) and providing the proportion of each type in these is not negligible that the contribution of the focal individual to the average perceived experience of each type becomes less and less important. In particular, we take it that $(1-\gamma_{[i]})N \gg 1$ and deduce that the average perceived aversiveness of type-1s in site i is approximately given as

$$I_{[i,1]} \approx (1-\gamma_{[i]})\frac{N}{n}L(r)H(t_1) + \gamma_{[i]}\frac{N}{n}L(r)H(0) \tag{6.2.8}$$

$$\approx (1-\gamma_{[i]})\mathcal{I}_1 + \gamma_{[i]}\mathcal{I}_2, \tag{6.2.9}$$

where we have used the shorthand notation $I_{1,2} = (N/n) \times L(r_1)H(t_{1,2})$ to denote the perceived aversiveness of a type-1/2 individual if it made up the entire locality. Extending the argument to type-2's, i.e. that $\gamma_{[i]}N \gg 1$ it is clear that under their average perceived aversiveness (6.2.7) is approximately equal to the quantity on the RHS of (6.2.8)

$$I_{[i,2]} \approx (1-\gamma_{[i]})\mathcal{I}_1 + \gamma_i\mathcal{I}_2 \approx I_{[i,1]}. \tag{6.2.10}$$

In particular, we establish that if a locality can carry a large enough number of prey and the proportions of each type within this are not negligible (i.e. we do not consider the $\gamma_{[i]} \to 0$ and $(1-\gamma_{[i]}) \to 0$ limits) then both types are perceived as equally aversive by the average predator visiting their site. Under this approximation the aversiveness of a prey individual depends on the relative proportion of type-1-to-type-2 individuals in its site and on the strategies that these play. We can therefore further simplify the subscript notation used so far so that it consists of just one index, $i = 1, ..., M$. We express the perceived aversiveness of a prey individual in locality i as

$$I_{[i]}(r_1, t_1) := (1-\gamma_{[i]})\frac{N}{n}L(r_1)H(t_1) + \gamma_{[i]}\frac{N}{n}L(r_1)H(0) \tag{6.2.11}$$

or more compactly as

$$I_{[i]} = (1-\gamma_{[i]})\mathcal{I}_1 + \gamma_{[i]}\mathcal{I}_2. \tag{6.2.12}$$

The significance of the square brackets in the subscript of $I_{[i]}$ should be pointed out here. While $I_{[i]}$ denotes the perceived aversiveness of a prey item in locality i consisting of a mix of $(1-\gamma_{[i]})N$ type-1 and $\gamma_{[i]}N$ type-2 individuals; $\mathcal{I}_{1/2}$ is the aversiveness of a prey residing in a population made up entirely of its type; it is not a site-dependent property is used mostly for notational convenience as in (6.2.12).

The following are true about the predator experience on the level of the site. Even though (i) individual experiences with type-1s are more aversive than those with type-2s ($H(t_1) > H(0)$) it is also true that (ii)

both classes of encounters are cognitively associated with a fixed appearance and that (iii) type-1s and type-2s are collectively perceived as equally aversive by the average predator visiting their site. This latter result is especially meaningful when considering Batesian mimicry systems and indeed one we would expect to recover in this representation of mimicry; if models and mimics were to be perceived as different and therefore to be attacked at different rates then this would not be consistent with the representation of mimicry systems in this chapter.

The proportion of type-1's and type-2's evaluated as averages over the sites is given as

$$1-\bar{\gamma}=\frac{\sum_{i=1}^{M}1-\gamma_{[i]}}{M}\ \text{ and }\ \bar{\gamma}:=\frac{\sum_{i=1}^{M}\gamma_{[i]}}{M}. \tag{6.2.13}$$

The spread of the proportions γ_i around this mean quantity is provided by the variance $\mathrm{Var}(\gamma)$, which is given as

$$\mathrm{Var}(\gamma):=\frac{\sum_{i=1}^{M}\left(\gamma_i-\bar{\gamma}\right)^2}{M}=\frac{\sum_{i=1}^{M}\gamma_i^2}{M}-\bar{\gamma}^{\,2}. \tag{6.2.14}$$

It is straightforward to show that $\mathrm{Var}(1-\gamma)=\mathrm{Var}(\gamma)$. From this we deduce the useful equality

$$\frac{\sum_{i=1}^{M}(1-\gamma_{[i]})^2}{M}=\mathrm{Var}(\gamma)+(1-\bar{\gamma})^2, \tag{6.2.15}$$

which we make use in the discussion that follows.

The aversive information as perceived by the average predator anywhere in the habitat is given by

$$\bar{I_1}=\frac{\sum_{i=1}^{M}(1-\gamma_{[i]})I_{[i]}N}{\sum_{i=1}^{M}(1-\gamma_{[i]})N} \tag{6.2.16}$$

$$=\frac{\mathcal{I}_1\sum_{i=1}^{M}(1-\gamma_{[i]})^2+\mathcal{I}_2\sum_{i=1}^{M}\gamma_i(1-\gamma_{[i]})}{\sum_{i=1}^{M}(1-\gamma_{[i]})}$$

$$=\frac{\mathcal{I}_1}{1-\bar{\gamma}}\times\frac{\sum_{i=1}^{M}(1-\gamma_i)^2}{M}+\frac{\mathcal{I}_2}{1-\bar{\gamma}}\times\frac{\sum_{i=1}^{M}\gamma_{[i]}}{M}-\frac{\mathcal{I}_2}{1-\bar{\gamma}}\times\frac{\sum_{i=1}^{M}\gamma_{[i]}^2}{M}$$

$$=\frac{\mathcal{I}_1}{1-\bar{\gamma}}\times\left[\mathrm{Var}(\gamma)+(1-\bar{\gamma})^2\right]+\mathcal{I}_2\times\frac{\bar{\gamma}}{1-\bar{\gamma}}-\frac{\mathcal{I}_2}{1-\bar{\gamma}}\times\left[\mathrm{Var}(\gamma)+\bar{\gamma}^{\,2}\right]$$

$$=\mathcal{I}_1\times(1-\bar{\gamma})+\mathcal{I}_2\times\bar{\gamma}+(\mathcal{I}_1-\mathcal{I}_2)\times\frac{\mathrm{Var}(\gamma)}{1-\bar{\gamma}}. \tag{6.2.17}$$

Similarly, we remark that the aversiveness of a type-2 individual as perceived by the average predator

anywhere in the habitat is

$$\bar{I_2} = \frac{\sum_{i=1}^{M} I_{[i]}\gamma_{[i]}N}{\sum_{i=1}^{M}\gamma_{[i]}N} \tag{6.2.18}$$

$$= \frac{\mathcal{I}_1\sum_{i=1}^{M}\gamma_{[i]}(1-\gamma_{[i]}) + \mathcal{I}_2\sum_{i=1}^{M}\gamma_{[i]}^2}{\sum_{i=1}^{M}\gamma_{[i]}}$$

$$= \mathcal{I}_1 - \frac{I_1}{\bar{\gamma}} \times \frac{\sum_{i=1}^{M}\gamma_i^2}{M} + \frac{\mathcal{I}_2}{\bar{\gamma}} \times \frac{\sum_{i=1}^{M}\gamma_i^2}{M}$$

$$= \mathcal{I}_1 - \frac{\mathcal{I}_1}{\bar{\gamma}} \times (\mathrm{Var}(\gamma) + \bar{\gamma}^2) + \frac{\mathcal{I}_2}{\gamma} \times (\mathrm{Var}(\gamma) + \bar{\gamma}^2)$$

$$= \mathcal{I}_1 \times (1-\bar{\gamma}) + \mathcal{I}_2 \times \bar{\gamma} - (\mathcal{I}_1 - \mathcal{I}_2) \times \frac{\mathrm{Var}(\gamma)}{\bar{\gamma}}. \tag{6.2.19}$$

Equalities (6.2.17) and (6.2.19) provide expressions for the aversivess of type-1 and type-2 individuals over the entire habitat in terms of an average base-line quantity $(1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2$ and an additional quantity, which controls the which is either added or subtracted from this. We write both these expressions here for ease of reference

$$\bar{I_1} = \mathcal{I}_1 \times (1-\bar{\gamma}) + \mathcal{I}_2 \times \bar{\gamma} + (\mathcal{I}_1 - \mathcal{I}_2) \times \frac{\mathrm{Var}(\gamma)}{1-\bar{\gamma}} \tag{6.2.20}$$

and

$$\bar{I_2} = \mathcal{I}_1 \times (1-\bar{\gamma}) + I_2 \times \bar{\gamma} - (\mathcal{I}_1 - \mathcal{I}_2) \times \frac{\mathrm{Var}(\gamma)}{\bar{\gamma}}. \tag{6.2.21}$$

From comparing (6.2.20) and (6.2.21) one concludes that $\bar{I_1} > \bar{I_2}$, which is a sensible conclusion. We would expect that since individual experiences with models are more aversive ($H(t_1) > H(0)$) that this is reflected in their overall perception by an average predator visiting the habitat.

Payoff

We now discuss and compare the fitness of models and mimics when this is perceived in the usual way (i.e. as a rate describing the number of offspring produced per life cycle). Drawing from our discussion on perceived aversiveness, we define this quantity on the level of the site and hence consider its average over the habitat in terms of a distribution of proportions. We should also remark that for purposes of avoiding excessive complexity, we only consider the case in which predation is the only source of death (i.e. $\lambda = 0$). We thus have that the fitness of a type-1 individual residing in site i is evaluated as

$$P_{[i,1]} := \frac{F_1(t_1)}{D(r_1)K_1(t_1)Q(I_{[i]})} \tag{6.2.22}$$

and likewise for a type-2 in that site this is

$$P_{[i,2]} := \frac{F_2(0)}{D(r_1)K_2(0)Q(I_{[i]})}, \tag{6.2.23}$$

where $I_{[i]} = (1-\gamma_{[i]})\mathcal{I}_1 + \gamma_{[i]}\mathcal{I}_2$ as in (6.2.12). We remark that expressions (6.2.20) and (6.2.21) are used to describe the payoff to an average type-1 and and average type-2 individual as

$$P_{\bar{1}} = \frac{F_1(t_1)}{D(r_1)K_1(t_1)Q(\bar{I}_1)} \quad \text{and} \quad P_{\bar{2}} = \frac{F_2(0)}{D(r_1)K_2(0)Q(\bar{I}_2)}. \tag{6.2.24}$$

The latter quantities are to be contrasted with the average payoff to a type-1 and type-2 individual (it is important to notice that the average behaviour of a system in general does not necessarily coincide with it's behaviour at the average).

Indeed, the procedure for evaluating the average type-1 and type-2 fitness is similar to that outlined in the lines following (6.2.16) and (6.2.18). For the first type we write

$$\begin{aligned} \bar{P}_1 &:= \frac{\sum_{i=1}^{M} P_{[i,1]}(1-\gamma_{[i]})N}{\sum_{i=1}^{M}(1-\gamma_{[i]})N} \\ &= \frac{F_1(t_1)}{D(r_1)K_1(t_1)} \frac{1}{1-\bar{\gamma}} \frac{1}{M} \sum_{i=1}^{M} \frac{1-\gamma_{[i]}}{Q(I_{[i]})} \end{aligned} \tag{6.2.25}$$

and similarly for the second type we have

$$\begin{aligned} \bar{P}_2 &:= \frac{\sum_{i=1}^{M} P_{[i,2]}\gamma_{[i]}N}{\sum_{i=1}^{M}\gamma_{[i]}N} \\ &= \frac{F_2(0)}{D(r_1)K_2(0)} \frac{1}{\bar{\gamma}} \frac{1}{M} \sum_{i=1}^{M} \frac{\gamma_{[i]}}{Q(I_{[i]})}. \end{aligned} \tag{6.2.26}$$

We should remark that most of the discussion in this section applies to populations that are finite and would therefore constitute a good starting point for considering real examples. But as we are also mathematically-minded we seek to find analytical results regarding the ecological and evolutionary stability (bi-stability) of Batesian mimicry systems in a manner that is efficient (and plausible). We proceed by re-introducing the Batesian complex described thus far in a manner that is more abstract and more convenient for the purposes of calculations. In particular, we treat the proportion of mimics as a random variable that is bound to a certain mass function and consider the limit as $N \to \infty$ in which the random variable is approximately continuous. The remainder of this chapter relies on the probability-theoretic approach discussed next.

6.3 A probability theory approach

In this section we consider the proportion of models to mimics as a discreet random assuming countably-many values on the unit interval (i.e. $\gamma = l/N$ with $l = 0, 1, ..., N-1, N$). We denote its probability mass function as $p(\gamma)$, its mean as $\bar{\gamma}$ and its associated cumulative mass function as $\mathcal{F}$. In the previous section we remarked that quantities such as the perceived aversiveness are approximately equal on the level of the site but different on average as they are assigned to types that exist in different proportions; compare for instance (6.2.12) with the average counterparts (6.2.20) and (6.2.21).

Continuous random variables

There is a clear analogy in the probabilistic picture, which we describe in generality here. Let $\mathcal{G}(\gamma)$ be some quantity whose value depends on the proportion of mimics. For the moment we suppress any dependence on the strategies to emphasize that $\mathcal{G}$ may be perceived as discreet random variable sharing the same mass function $p(\gamma)$ as γ itself. As is clear from the previous section, we are especially interested in the average behaviour of the system and therefore on the *expected value* of such quantities.

An intuitive description of the expected value is that of an average weighted over the mass function

$$\sum_{\gamma:p(\gamma)>0} \mathcal{G}(\gamma)p(\gamma), \tag{6.3.1}$$

where the latter is clearly equivalent to

$$\sum_{l=0}^{N} \mathcal{G}(l/N)\left[\mathcal{F}(l/N) - \mathcal{F}((l-1)/N)\right]. \tag{6.3.2}$$

We define the *partition* of the closed interval $[a,b] \subset \mathbb{R}$ as a finite subset of $[a,b]$ containing the end-points a and b (see definition provided in section 5.2 of Salas et al., 2007). Namely, if

$$\mathcal{V} = \{x_0, x_1, ...x_v\} \tag{6.3.3}$$

is a partition of $[a,b]$, then it breaks up $[a,b]$ into v subintervals of lengths $\Delta x_1, \Delta x_2, ..., \Delta x_v$. The *norm* $||\cdot||$ of partition (6.3.3) is defined as the length of the largest subinterval, i.e. $||\mathcal{V}|| := max(\Delta x_v)$

The set $\{0, 1/N, 2/N, ..., 1\}$ is a partition of the unit interval $[0,1]$ and in the limit as $N \to \infty$ the norm of this partition tends to zero. In this limit the sum in (6.3.2) and can be approximated by the Stieltje integral, which in turn provides a more general definition for the expectation

$$\lim_{N\to\infty} \sum_{l=0}^{N} \mathcal{G}(l/N)\left[\mathcal{F}(l/N) - \mathcal{F}((l-1)/N)\right] = \int_0^1 \mathcal{G}(\gamma)d\mathcal{F}(\gamma) =: \bar{\mathcal{G}}. \tag{6.3.4}$$

We should also remark that in this limit the cumulative mass function $\mathcal{F}$ is approximately continuous and provided this is almost everywhere $\mathcal{C}^2$ we may apply linearisation to the quantity in the integrand $d\mathcal{F}(\gamma) \approx f(\gamma)d\gamma$, where $f(\gamma) = \mathcal{F}'(\gamma)$ is the density function of the (continuous) random variable γ.

In this continuum limit we therefore express the expectation of the quantities as

$$\bar{\mathcal{G}} = \int_0^1 \mathcal{G}(\gamma)f(\gamma)d\gamma. \tag{6.3.5}$$

As we have seen in the previous subsection, certain quantities such as the perceived aversiveness or indeed measures of fitness are type-specific, in the sense that they describe individuals of one type or the other uniquely; one can contrast (6.2.22) and (6.2.23) for instance. Indeed, such a type-1 quantity can be perceived as a continuous random variable $\mathcal{G}_1$ with distribution

$$f_1(\gamma) := \frac{(1-\gamma)f(\gamma)}{\int_0^1 (1-\gamma)f(\gamma)d\gamma} = \frac{(1-\gamma)f(\gamma)}{1-\bar{\gamma}}, \tag{6.3.6}$$

and similarly that the type-2 quantity $\mathcal{G}_2$ follows the distribution

$$f_2(\gamma) := \frac{\gamma f(\gamma)}{\bar{\gamma}}. \tag{6.3.7}$$

In accordance with (6.3.5) it follows that the expected values of such type-specific quantities are provided as

$$\bar{\mathcal{G}}_1 = \frac{\int_0^1 \mathcal{G}_1(\gamma)(1-\gamma)f(\gamma)d\gamma}{1-\bar{\gamma}} \tag{6.3.8}$$

and

$$\bar{\mathcal{G}}_2 = \frac{\int_0^1 \mathcal{G}_2(\gamma)\gamma f(\gamma)d\gamma}{\bar{\gamma}}. \tag{6.3.9}$$

Aversive information

We can use expressions (6.3.8) and (6.3.9) to determine the expected values for the aversiveness and fitness. In the previous section it was shown that in the $N \to \infty$ limit the perceived aversiveness in any given site depends only on the mimic/model proportion in that site, such that either type is perceived as having the same aversiveness by the average predator visiting that site. By analogy and by maintaining these assumptions we define the perceived aversiveness for fixed model and mimic strategies (r_1, t_1) and $(r_1, 0)$ as the random variable

$$I(\gamma) = (1-\gamma)\mathcal{I}_1 + \gamma\mathcal{I}_2. \tag{6.3.10}$$

We should remark that in spite of (6.3.10) the perceived aversiveness is by construction a type-specific quantity, such that among type-1s follows the distribution $f_1(\gamma)$ and among type-2s the distribution $f_2(\gamma)$. The expected type-1 aversiveness is given as

$$\begin{aligned}
\bar{I}_1 &= \frac{\int_0^1 I(\gamma)(1-\gamma)f(\gamma)d\gamma}{1-\bar{\gamma}} \\
&= \mathcal{I}_1 \frac{\int_0^1 (1-\gamma)^2 f(\gamma)d\gamma}{1-\bar{\gamma}} + \mathcal{I}_2 \frac{\int_0^1 \gamma(1-\gamma)f(\gamma)d\gamma}{1-\bar{\gamma}} \\
&= (1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2 + (\mathcal{I}_1 - \mathcal{I}_2)\frac{\mathrm{Var}(\gamma)}{1-\bar{\gamma}}
\end{aligned} \tag{6.3.11}$$

and indeed the expected type-2 aversiveness is

$$\begin{aligned}
\bar{I}_2 &= \frac{\int_0^1 I(\gamma)\gamma f(\gamma)d\gamma}{\bar{\gamma}} \\
&= (1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2 - (\mathcal{I}_1 - \mathcal{I}_2)\frac{\mathrm{Var}(\gamma)}{\bar{\gamma}},
\end{aligned} \tag{6.3.12}$$

both of which validate our findings of (6.2.20) and (6.2.21) from the previous section.

Payoff

The fitness is also a type-specific quantity and by extension, with $\lambda = 0$ the type-1 fitness is defined as the variable

$$P_1(\gamma) = \frac{F_1(t_1)}{D(r_1)K_1(t_1)Q(I)} \tag{6.3.13}$$

that is randomly distributed with probability density $f_1(\gamma)$ of expression (6.3.6). Likewise and for some choice of strategies the type-2 fitness is defined as the random variable

$$P_2(\gamma) = \frac{F_2(0)}{D(r_1)K_2(0)Q(I)} \tag{6.3.14}$$

with density function $f_2(\gamma)$ as in (6.3.7). We remark that while for given model and mimic strategy choices, expressions (6.3.13) and (6.3.14) can be perceived as random variables, it is also the case that for some observation of γ these may also be perceived as $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ functions - note that the same applies to I in (6.3.10). Indeed, according to (6.3.8) and (6.3.9) the expected values of fitness for either type are

$$\bar{P_1}(r_1, t_1; r_1, 0) = \frac{F_1(t_1)}{D(r_1)K_1(t_1)} \frac{1}{1-\bar{\gamma}} \int_0^1 \frac{1-\gamma}{Q(I)} f(\gamma) d\gamma \tag{6.3.15}$$

and

$$\bar{P_2}(r_1, t_1; r_1, 0) = \frac{F_2(0)}{D(r_1)K_2(0)} \frac{1}{\bar{\gamma}} \int_0^1 \frac{\gamma}{Q(I)} f(\gamma) d\gamma. \tag{6.3.16}$$

The beta distribution & payoff

The process of sampling prey at random from within their localities (with replacement, although this is less important under the $N \gg 1$ assumption) consists of repeating Bernoulli trials, where "success" is understood as the event that a type-2 individual is encountered. The random variable defined as the number of type-2 individuals encountered follows the binomial distribution with probability γ of success. The usual conjugate prior for this random variable is the *beta distribution* with *hyper-parameters* $\alpha, \beta \in \mathbb{R}^{>0}$, both of which are shape parameters (the reader is encouraged to consult chapter 5 of Ross, 2019 for introductory details into the beta distribution and chapter 25 of Johnson et al., 1995 for a more formal presentation on the topic). The distribution for γ is given as

$$f(\gamma) = \frac{\gamma^{\alpha-1}(1-\gamma)^{\beta-1}}{B(\alpha, \beta)}, \tag{6.3.17}$$

where the term $B(\alpha, \gamma)$ is known as the *beta function* and is defined in terms of α and β as

$$B(\alpha, \beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx. \tag{6.3.18}$$

We remark that if $\alpha < \beta$ the distribution is skewed left, while if $\alpha > \beta$ the distribution is skewed right. If $\alpha = \beta$ the distribution is dispersed equally around $\bar{\gamma}$ and in the special case $\alpha = \beta = 1$ the distribution is uniform. Equalities

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad \text{and} \quad \Gamma(x+1) = x\Gamma(x), \tag{6.3.19}$$

where $\Gamma(x)$ is the *gamma function* can be used to show that the expectation $\bar{\gamma}$ and variance $\mathrm{Var}(\gamma)$ of (6.3.17) amount to

$$\bar{\gamma} = \frac{\alpha}{\alpha + \beta} \tag{6.3.20}$$

and

$$\mathrm{Var}(\gamma) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \tag{6.3.21}$$

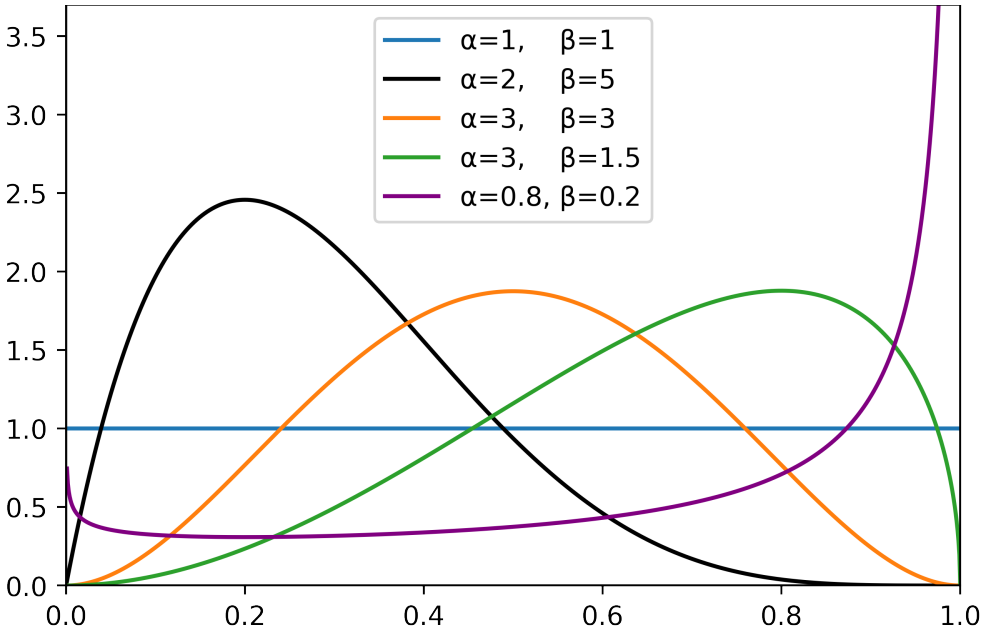

Figure 6.1: Plots of beta probability density functions with shape parameters as specified in the legend. By construction in our context the x-axis represents the proportion of mimics γ over the population. The black curve with parameters $\alpha = 2, \beta = 5$ is used in section 6.6 to generate a working example of a Batesian mimicry complex that is eco-evolutionarily stable.

In light of (6.3.17) the average payoffs of type-1 and type-2 can be expressed as

$$\bar{P}_1 = \frac{F_1(t_1)}{D(r_1)K_1(t_1)} \frac{1}{1-\bar{\gamma}} \frac{1}{B(\alpha,\beta)} \int_0^1 \frac{\gamma^{\alpha-1}(1-\gamma)^{\beta}}{Q(I)} d\gamma \tag{6.3.22}$$

and

$$\bar{P}_2 = \frac{F_2(0)}{D(r_1)K_2(0)} \frac{1}{\bar{\gamma}} \frac{1}{B(\alpha,\beta)} \int_0^1 \frac{\gamma^{\alpha}(1-\gamma)^{\beta-1}}{Q(I)} d\gamma \tag{6.3.23}$$

We define parameter δ as the sum of the shape parameters $\delta := \alpha + \beta$, which allows us to re-express the shape parameters in terms of the mean $\bar{\gamma}$ and δ

$$\alpha = \delta\bar{\gamma} \quad \text{and} \quad \beta = \delta(1-\bar{\gamma}). \tag{6.3.24}$$

Expressing the variance in terms of the new substituted variables $\bar{\gamma}$ and δ allows us to conclude that δ is indeed an honest measure of the closeness of spread of the distribution around the mean. From (6.3.21) we have

$$\mathrm{Var}(\gamma) = \frac{\bar{\gamma}(1-\bar{\gamma})}{\delta+1}. \tag{6.3.25}$$

Carrying through the variable substitution of (6.3.24) into the density (6.3.17) and using notation $\hat{\cdot}$ to distinguish the substituted quantities we write

$$\hat{f}(\gamma) = \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{\hat{B}(\bar{\gamma},\delta)}, \tag{6.3.26}$$

with

$$\hat{B}(\bar{\gamma},\delta) := B(\delta\bar{\gamma},\delta(1-\bar{\gamma})) = \int_0^1 x^{\delta\bar{\gamma}-1}(1-x)^{\delta(1-\bar{\gamma})-1}dx. \tag{6.3.27}$$

Following the above discussion it is straightforward to establish that type-1 quantities have density functions

$$f_1(\gamma) = \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{(1-\bar{\gamma})\hat{B}(\bar{\gamma},\delta)} \text{ and } f_2(\gamma) = \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{\bar{\gamma}\hat{B}(\bar{\gamma},\delta)}. \tag{6.3.28}$$

It should be noted that (6.3.11) and (6.3.12) transform in the straightforward way, such that

$$\bar{I}_1 = (1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2 + \frac{\bar{\gamma}}{1+\delta}(\mathcal{I}_1 - \mathcal{I}_2) \tag{6.3.29}$$

and

$$\bar{I}_2 = (1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2 - \frac{1-\bar{\gamma}}{1+\delta}(\mathcal{I}_1 - \mathcal{I}_2). \tag{6.3.30}$$

From the latter it is clear that for some fixed distribution mean $\bar{\gamma}$ the larger the parameter δ the smaller the difference between $\bar{I}_1$ and $\bar{I}_2$, indeed we have

$$\lim_{\delta\to\infty}\bar{I}_1 = \lim_{\delta\to\infty}\bar{I}_2 = (1-\bar{\gamma})\mathcal{I}_1 + \bar{\gamma}\mathcal{I}_2. \tag{6.3.31}$$

Conversely, the smaller δ is (for fixed mean $\bar{\gamma}$) the larger the difference $\bar{I}_1 - \bar{I}_2$ between the average perceived aversiveness of the two types. We should remark that while it is possible to consider the limit $\delta \to \infty$, the limit $\delta \to 0$ would imply that the shape parameters vanish so that $\alpha \to 0$ and $\beta \to 0$ where the Beta distribution diverges.

Finally, we carry through the substituted variables of (6.3.24) into the expected payoff to arrive at the more convenient forms

$$\bar{P}_1(r_1,t_1;r_1,0) = \frac{F_1(t_1)}{D(r_1)K_1(t_1)}\frac{1}{1-\bar{\gamma}}\frac{1}{\hat{B}(\bar{\gamma},\delta)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)}d\gamma \tag{6.3.32}$$

and

$$\bar{P}_2(r_1,t_1;r_1,0) = \frac{F_2(0)}{D(r_1)K_2(0)}\frac{1}{\bar{\gamma}}\frac{1}{\hat{B}(\bar{\gamma},\delta)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma. \tag{6.3.33}$$

We have now established two key integral forms (6.3.32) and (6.3.33) for the expected type-1 and type-2 payoffs, which apply in the limit as $N \to \infty$ where the proportion of mimics to models can be approximated as a continuous random variable. In the sections that follow, we discuss eco-evolutionary (similarly to how this was done for chapter 5) and determine the conditions under which two types forming a Batesian mimicry complex can co-exist. In order to achieve this we introduce two separate notions of stability, which act on separate time-scales: ecological (shorter) and evolutionary (longer); for the latter we extend the ideas that were introduced in the single species description.

6.4 Ecological stability

In this section we build on the results of the previous section, which assume that the proportion of mimics to models is an approximately continuous, beta-distributed random variable and develop the conditions for ecological stability. These conditions are summarised by the requirement that (on average) models and mimics produce the same number of offspring per life-cycle (i.e. achieve equal fitness) and that this equilibrium is stable in small perturbations of their proportions. We should remark that in this section the ecological stability conditions are discussed in absence of mutation and although necessary are as such not sufficient to describe co-existence on the (longer) evolutionary timescales. To complete the picture, we introduce mutation separately in the next section and extend the notion of evolutionary stability from its definition the single-species case to Batesian mimicry complexes that have been the subject of this chapter. Following this, ecological and evolutionary stability are considered jointly

We say that the population is in ecological equilibrium if on average models and mimics produce the same number of offspring per life-cycle. Such a situation suggests that on average neither population is increasing or decreasing compared with the other and that these must co-exist in some equilibrium proportion $(1-\bar{\gamma}^*, \gamma^*)$. This is provided by

$$\bar{P}_1(r_1, t_1; r_1, 0) = \bar{P}_2(r_1, t_1; r_1, 0). \tag{6.4.1}$$

Substitution of (6.3.32) and (6.3.33) suggest that $\bar{\gamma}^*$ is defined through equality

$$\frac{F_1(t_1)}{K_1(t_1)} \frac{1}{1-\bar{\gamma}^*} \int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)} d\gamma = \frac{F_2(0)}{K_2(0)} \frac{1}{\bar{\gamma}^*} \int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)} d\gamma. \tag{6.4.2}$$

The situation in (6.4.2) is stable under perturbations of the mimic/model proportions $(\gamma^*, 1-\gamma^*)$ if in addition to (6.4.2) we impose

$$\left[\partial_{\bar{\gamma}} \bar{P}_2 - \partial_{\bar{\gamma}} \bar{P}_1\right] |_{\bar{\gamma}=\bar{\gamma}^*} < 0. \tag{6.4.3}$$

The latter is equivalent to inequality

$$\hat{B}(\bar{\gamma}, \delta) D(r_1) \partial_{\bar{\gamma}} \bar{P}_1 - \hat{B}(\bar{\gamma}, \delta) D(r_1) \partial_{\bar{\gamma}} \bar{P}_2 > 0, \tag{6.4.4}$$

which will prove more convenient for the body of calculations that follow. Indeed, substitution of (6.3.32) into the first term of (6.4.4) amounts to

$$\hat{B}(\bar{\gamma}, \delta) D(r_1) \partial_{\bar{\gamma}} \bar{P}_1 =$$

$$= \frac{1}{(1-\bar{\gamma})^2} \frac{F_1(t_1)}{K_1(t_1)} \int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)} d\gamma - \frac{1}{1-\bar{\gamma}} \frac{F_1(t_1)}{K_1(t_1)} \frac{\partial_{\bar{\gamma}} \hat{B}(\bar{\gamma}, \delta)}{\hat{B}(\bar{\gamma}, \delta)} \int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)} d\gamma$$

$$+ \frac{\delta}{1-\bar{\gamma}} \frac{F_1(t_1)}{K_1(t_1)} \int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})} \log\left(\frac{\gamma}{1-\gamma}\right)}{Q(I)} d\gamma \tag{6.4.5}$$

while substitution of (6.3.33) for the second term in (6.4.4) gives us

$$-\hat{B}(\bar{\gamma},\delta)D(r_1)\partial_{\bar{\gamma}}\bar{P}_2 =$$

$$= -\frac{1}{\bar{\gamma}^2}\frac{F_2(0)}{K_2(0)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma - \frac{1}{\bar{\gamma}}\frac{F_2(0)}{K_2(0)}\frac{\partial_{\bar{\gamma}}\hat{B}(\bar{\gamma},\delta)}{\hat{B}(\bar{\gamma},\delta)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma$$

$$+\frac{\delta}{\bar{\gamma}}\frac{F_2(0)}{K_2(0)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}\log\left(\frac{\gamma}{1-\gamma}\right)}{Q(I)}d\gamma > 0. \tag{6.4.6}$$

From scaling the equilibrium condition by $\partial_{\bar{\gamma}}\hat{B}(\bar{\gamma},\delta)$ we observe that the second terms in (6.4.5) and (6.4.6) are identical. From the same condition we also express the first terms in (6.4.5) and (6.4.6) in terms of a single common integral. In particular, condition (6.4.4) now amounts to

$$\frac{1}{\bar{\gamma}^2(1-\bar{\gamma})}\frac{F_2(0)}{K_2(0)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma$$

$$+\frac{\delta}{1-\bar{\gamma}}\frac{F_1(t_1)}{K_1(t_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}\log\left(\frac{\gamma}{1-\gamma}\right)}{Q(I)}d\gamma$$

$$-\frac{\delta}{\bar{\gamma}}\frac{F_2(0)}{K_2(0)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}\log\left(\frac{\gamma}{1-\gamma}\right)}{Q(I)}d\gamma > 0. \tag{6.4.7}$$

For the subset of Beta distributions with some fixed mean $\bar{\gamma}$ the equal payoffs condition in (6.4.2) specifies a curve in the strategy space. Intersections of this curve with subregions identified by inequality (6.4.7) contains strategies that are stable under small perturbations in the mimic and model proportions. In the majority of the analysis that follows we tend to think of the mean $\bar{\gamma}$ as a variable and not as a parameter such that the equal payoffs condition (6.4.2) can be thought to describe a surface in the $(r_1, t_1, \bar{\gamma})$-space and inequality (6.4.7) as specifying regions of the surface where the condition for equal payoffs is consistent under small perturbations in the average mimic/model proportions.

In the following section we introduce local mutation to the model and mimic strategies and observe a similar effect, namely that the complex is locally evolutionarily stable when the model and mimic strategies and associated mean proportions are drawn from sections of a surface in the $(r_1, t_1\bar{\gamma})$-space. Following ESS analysis we unify the two pictures and hence discuss explicit examples of systems that are eco-evolutionarily stable (i.e. satisfy the conditions for stability mentioned in this and the following chapter).

6.5 Evolutionary stability

In this section we extend the notion of co-existence introduced in the previous section to account for mutation. In addition to the stability conditions discussed therein, we now require that the model and mimic populations are individually uninvadable by mutations occurring locally to their type. A Batesian mimicry complex is said to be eco-evolutionarily stable if the two types on average achieve equal payoffs in the sense of (6.4.1) - and this condition is consistent under (6.4.3) and in addition, each type is uninvadable by mutations occurring locally to its type. As in the previous section, we work in the $N \to \infty$ limit in which the mimic/model

proportion is effectively a continuous random variable.

For some given proportion γ, the model and mimic fitness are

$$P_1(r_1,t_1;r_1,0)=\frac{F_1(t_1)}{D(r_1)K_1(t_1)Q(I)} \tag{6.5.1}$$

and

$$P_2(r_1,t_1;r_1,0)=\frac{F_2(0)}{D(r_1)K_2(0)Q(I)} \tag{6.5.2}$$

and follow beta distributions, as discussed earlier. The payoff associated to a mutant playing strategy $(r,t)\in[r_1-\delta r,r_1+\delta r]\times[t_1-\delta t,t_1+\delta t]\setminus(r_1,t_1)$ is given by

$$P_1^\dagger(r,t;r_1,t_1;r_1,0)=\frac{F_1(t)}{D(r)K_1(t)Q(I_1^\dagger)} \tag{6.5.3}$$

where

$$I_1^\dagger=a(1-\gamma)\mathcal{I}+(1-a)(1-\gamma)\mathcal{I}_1S(|r-r_1|)+\gamma\mathcal{I}_2S(|r-r_1|) \tag{6.5.4}$$

describes its aversiveness as perceived by the average predator. Likewise, the payoff to a mutant playing strategy $(r,t)\in[r_1-\delta r,r_1+\delta r]\times[0,\delta t]\setminus(r_1,0)$ local to the mimic strategy is given by

$$P_2^\dagger(r,t;r_1,t_1;r_1,0)=\frac{F_2(t)}{D(r)K_2(t)Q(I_2^\dagger)}, \tag{6.5.5}$$

where

$$I_2^\dagger=a\gamma\mathcal{I}+(1-a)\gamma\mathcal{I}_2S(|r-r_1|)+(1-\gamma)\mathcal{I}_1S(|r-r_1|) \tag{6.5.6}$$

denotes its aversiveness as perceived by the average predator. We remark that the † superscript is used to distinguish mutant quantities from their resident (model/mimic) counterparts, which are (as usual) indicated with the subscripts $_1$ and $_2$. In addition, since the functional forms F,D,K,Q,L,H and S (sufficiently near the origin) are $\mathcal{C}^l$ with $l\geq 2$ and quantities $I_1^\dagger$ and $I_2^\dagger$ depend on $|r-r_1|$ it follows that the mutant fitness functions $P_1^\dagger$ and $P_2^\dagger$ are almost everywhere $\mathcal{C}^l$ with $l\geq 2$ except at $r=r_1$ (where they are not differentiable with respect to r). It should be clear that when the mutant plays the model strategy (r_1,t_1) or the mimic strategy $(r_1,0)$ that both expressions (6.5.4) and (6.5.6) both amount to the same quantity, which we denote I. This is

$$I=(1-\gamma)\mathcal{I}_1+\gamma\mathcal{I}_2. \tag{6.5.7}$$

We should remark that expressions (6.5.4) and (6.5.6) are consistent with our intuition that a is the average local relatedness on the level of the type, such that a level of local relatedness a corresponds to a total of $a(1-\gamma)N$ type-1 or $a\gamma N$ type-2 mutants in a site containing mutants. The reason for this is that usually the competing mimic is of a different species.

Notice that there are three pairs of strategies on the arguments of the quantities on the LHSs of (6.5.3) and (6.5.5), which account for the mutant, model and mimic strategies. To retain consistency, we remark the following, namely that while the model and mimic payoffs in (6.5.8) and (6.5.9) depend only on the model and mimic strategies (indeed these are $\mathbb{R}^2\times\mathbb{R}^2\to\mathbb{R}$ functions), while their mutant counterparts (†) naturally depend also on the mutant strategy which is defined in the local vicinity of either of the first two. For consistency we make explicit the fact that the (expected) mutant payoffs of (6.5.3) and (6.5.5) when evaluated at the model/mimic strategy return the original model/mimic fitness.

As discussed in the previous section, we tend to consider the limit as $N \to \infty$ where the proportion of mimics in a given site/observation can be approximated as a continuous random variable with at least $\mathcal{C}^1$ density function $f(\gamma)$ - it should also be remarked that although f is kept general at this stage, in later stages $f(\gamma)$ will represent some choice of Beta distribution. Following the reasoning of the previous section it should be clear that the quantities in (6.5.3) and (6.5.5) are also random variables sharing this distribution. Their expected values over the distribution are

$$\bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0) = \frac{F_1(t)}{D(r)K_1(t)}\frac{1}{1-\bar{\gamma}}\int_0^1 \frac{(1-\gamma)f(\gamma)}{Q(I_1^\dagger)}d\gamma \tag{6.5.8}$$

and

$$\bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0) = \frac{F_2(t)}{D(r)K_2(t)}\frac{1}{\bar{\gamma}}\int_0^1 \frac{\gamma f(\gamma)}{Q(I_2^\dagger)}d\gamma. \tag{6.5.9}$$

We emphasize that while the mutant fitness functions $P_1^\dagger$ and $P_2^\dagger$ provided in (6.5.3) and (6.5.5) are almost everywhere $\mathcal{C}^l$ with $l \geq 2$ in the vicinity of the model and mimic strategies (r_1,t_1) and $(r_1,0)$ these are not differentiable with respect to the mutant conspicuousness at $r = r_1$. It should be clear to the reader that the same claims are valid for the averages of these quantities $\bar{P}_1$ and $\bar{P}_2$ given in (6.5.8) and (6.5.9). In the definition that follows - Definition 6.5.1 - we explain what it means for each type to be locally evolutionarily stable.

Definition 6.5.1. *Assume that in a Batesian mimicry complex in which the aversive models play (r_1,t_1) with $r_1 > 0$ and $t_1 > t_c$ in proportion $1-\gamma$, where γ is a continuous, Beta-distributed random variable such that $\gamma \sim Beta(\delta\bar{\gamma},\delta(1-\bar{\gamma}))$ and the mimics play $(r_1,0)$ in proportion $1-\gamma$. In addition, assume that the (average) fitness received by mutants playing strategy (r,t) local to the model/mimic strategy are denoted $\bar{P}_1^\dagger/\bar{P}_2^\dagger$. We say that:*

[i] The models are locally evolutionarily stable if (on average) they receive higher fitness when interacting with the model strategy than do the mutants (that are local to the model strategy) when interacting with the model strategy. That is, we say that the model strategy is locally evolutionarily stable if

$$\bar{P}_1^\dagger(r_1,t_1;r_1,t_1;r_1,0) > \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0) \text{ for all } (r,t) \in [r_1-\delta r, r_1+\delta r]\times[t_1-\delta t, t_1+\delta t]\setminus(r_1,t_1). \tag{6.5.10}$$

[ii] The mimics are locally evolutionarily stable if (on average) they receive higher fitness when interacting with the mimic strategy than do the mutants (that are local to the mimic strategy) when interacting with the mimic strategy. That is, we say that the mimic strategy is locally evolutionarily stable if

$$\bar{P}_2^\dagger(r_1,0;r_1,t_1;r_1,0) > \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0) \text{ for all } (r,t) \in [r_1-\delta r, r_1+\delta r]\times[0,\delta t]\setminus(r_1,0). \tag{6.5.11}$$

Finally, we say that the complex is evolutionarily stable if both the models and the mimics are locally evolutionarily stable in the sense of [i] and [ii] as above.

The above definition for local ESS follows as a direct extension of the original definition (Definition 2.3.1) provided in chapter 2. From this it is clear that if the (average) type-1 mutant fitness admits a strict local maximum at the model strategy and the (average) type-2 mutant fitness to admits a strict local maximum at the mimic strategy then the Batesian mimicry complex is locally evolutionarily stable. In the section that follows it is clarified that we require both the model and the mimic to be locally evolutionarily stable (in addition to being ecologically stable - see previous section) for the complex to be considered stable as a

collective. We make use of the following theorem (Theorem 6.5.2) in order to determine whether a model or mimic strategy is local ESS, which is a direct extension of the theorem (Theorem 2.3.2) used in chapter 2 to determine whether a resident strategy is a local ESS. We state and prove this theorem presently.

Theorem 6.5.2. *Assume that in a Batesian mimicry complex the aversive models play* (r_1, t_1) *with* $r_1 > 0$ *and* $t_1 > t_c$ *in proportion* $1-\gamma$*, where* γ *is a continuous, beta-distributed random variable such that* $\gamma \sim Beta(\delta\bar{\gamma}, \delta(1-\bar{\gamma}))$ *and the mimics play* $(r_1, 0)$ *in proportion* $1-\gamma$*. Assume that the average fitnesses received by mutants playing strategy* (r,t) *local to the model/mimic value are denoted* $\bar{P}_1^\dagger/\bar{P}_2^\dagger$ *and given through* (6.5.8)/(6.5.9)*. The quantities* $\bar{P}_1^\dagger$ *and* $\bar{P}_2^\dagger$ *are almost everywhere* $\mathcal{C}^l$ *with* $l \geq 2$ *except at* $r = r_1$ *where they are not* r*-differentiable but are continuous at that value. It then follows that if:*
$[i]$

$$\partial_t \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1} = 0, \tag{6.5.12}$$

$$\partial_{tt} \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1} < 0 \tag{6.5.13}$$

$$\overleftarrow{\partial}_r \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1} > 0 \ \textit{and} \tag{6.5.14}$$

$$\overrightarrow{\partial}_r \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1} < 0 \tag{6.5.15}$$

then the model strategy (r_1, t_1) *is a local ESS.*
In addition, if
$[ii]$

$$\overrightarrow{\partial}_t \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0} < 0 \tag{6.5.16}$$

$$\overleftarrow{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0} > 0 \ \textit{and} \tag{6.5.17}$$

$$\overrightarrow{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0} < 0, \tag{6.5.18}$$

then the mimic strategy $(r_1, 0)$ *is a local ESS.*

Proof. We show that inequalities (6.5.12), (6.5.13), (6.5.14) and (6.5.15) in $[i]$ lead to local ESS for the model in the sense of (6.5.10) in Definition 6.5.1. First, we assume that mutation is present in the model strategy (r_1, t_1) and express the mutant traits in terms of spherical coordinates, such that

$$(x,\phi) \to (r,t): \ \ r = r_1 + x\cos\phi \ \text{ and } \ t = t_1 + x\sin\phi. \tag{6.5.19}$$

Hence, we express the average fitness $\bar{\mathcal{P}}_1^\dagger$ received by such a mutant in terms of the transformed coordinates x, ϕ so that

$$\bar{\mathcal{P}}_1^\dagger : \mathbb{R}^{\geq 0} \times [0, 2\pi) \to \mathbb{R}^{\geq 0}: \ \ \bar{\mathcal{P}}_1^\dagger(x,\phi) := \bar{P}_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_1, 0). \tag{6.5.20}$$

From the latter it is clear that the desired inequality (6.5.10) in Definition 6.5.1 amounts to showing

$$\bar{\mathcal{P}}_1^\dagger(x,\phi) - \bar{\mathcal{P}}_1^\dagger(0,\phi) < 0 \ \text{ for all } \ \phi \in [0, 2\pi). \tag{6.5.21}$$

We proceed to showing this for cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2, 3\pi/2$, $\phi \in (\pi/2, \pi)$, $\phi = \pi$, $\phi \in (\pi, 3\pi/2)$ and $\phi \in (3\pi/2, 2\pi)$ individually.

If $\phi = 0$ mutation is solely along the r-direction so that

$$\begin{aligned}\bar{\mathcal{P}}_1^\dagger(x, \phi = 0) - \bar{\mathcal{P}}_1^\dagger(0, \phi = 0) &= \bar{P}_1^\dagger(r, t_1; r_1, t_1; r_1, 0) - \bar{P}_1^\dagger(r_1, t_1; r_1, t_1; r_1, 0) \\ &\approx \underbrace{(r - r_1)}_{>0} \times \underbrace{\overrightarrow{\partial}_r \bar{\mathcal{P}}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{<0} < 0. \end{aligned} \tag{6.5.22}$$

If $\phi \in (0, \pi/2)$ we have

$$\begin{aligned}&\bar{\mathcal{P}}_1^\dagger(x, \phi = 0) - \bar{\mathcal{P}}_1^\dagger(0, \phi = 0) \approx x\partial_x \bar{\mathcal{P}}_1^\dagger(x, \phi)|_{x=0, \phi \in (0, \pi/2)} \\ &= x\partial_x \bar{P}_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_1, 0)|_{x=0, \phi \in (0, \pi/2)} \\ &= \underbrace{x\cos\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_r \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{=0} < 0. \end{aligned} \tag{6.5.23}$$

If $\phi = \pi/2$ or $\phi = 3\pi/2$ mutation is along the t-direction so that $x = t - t_1$ and thus

$$\begin{aligned}&\bar{\mathcal{P}}_1^\dagger(x, \phi) - \bar{\mathcal{P}}_1^\dagger(0, \phi) = \bar{P}_1^\dagger(r_1, t; r_1, t_1; r_1, 0) - \bar{P}_1^\dagger(r_1, t_1; r_1, t_1; r_1, 0) \\ &= (t - t_1) \times \underbrace{\partial_t \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{=0} + \frac{1}{2} \underbrace{(t - t_1)^2}_{>0} \times \underbrace{\partial_{tt} \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{<0} < 0. \end{aligned} \tag{6.5.24}$$

If $\phi \in (\pi/2, \pi)$ the (average) incremental difference between the mutant and the model fitness is given by

$$\begin{aligned}&\bar{\mathcal{P}}_1^\dagger(x, \phi) - \bar{\mathcal{P}}_1^\dagger(0, \phi) \approx x\partial_x \bar{\mathcal{P}}_1^\dagger(x, \phi)|_{x=0, \phi \in (\pi/2, \pi)} \\ &= x\partial_x \bar{P}_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_1, 0)|_{x=0, \phi \in (\pi/2, \pi)} \\ &= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{>0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\partial_t \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{=0} < 0. \end{aligned} \tag{6.5.25}$$

If $\phi = \pi$ we have mutation along the r-direction so that $x = r - r_1 < 0$ and thus

$$\begin{aligned}\bar{\mathcal{P}}_1^\dagger(x, \phi = \pi) - \bar{\mathcal{P}}_1^\dagger(0, \phi = \pi) =& \bar{P}_1^\dagger(r, t_1; r_1, t_1; r_1, 0) - \bar{P}_1^\dagger(r_1, t_1; r_1, t_1; r_1, 0) \\ =& \underbrace{(r - r_1)}_{<0} \times \underbrace{\overleftarrow{\partial}_r \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{>0} < 0. \end{aligned} \tag{6.5.26}$$

If $\phi \in (\pi, 3\pi/2)$

$$\begin{aligned}&\bar{\mathcal{P}}_1^\dagger(x, \phi) - \bar{\mathcal{P}}_1^\dagger(0, \phi) \approx x\partial_x \bar{\mathcal{P}}_1^\dagger(x, \phi)|_{x=0, \phi \in (\pi, 3\pi/2)} \\ &= x\partial_x \bar{P}_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_1, 0)|_{x=0, \phi \in (\pi, 3\pi/2)} \\ &= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{>0} + x\sin\phi \times \underbrace{\partial_t \bar{P}_1^\dagger(r, t; r_1, t_1; r_1, 0)|_{r=r_1, t=t_1}}_{=0} < 0. \end{aligned} \tag{6.5.27}$$

Finally, if $\phi \in (3\pi/2, 2\pi)$ it follows that

$$\bar{\mathcal{P}}_1^\dagger(x,\phi) - \bar{\mathcal{P}}_1^\dagger(0,\phi) \approx x\partial_x \bar{\mathcal{P}}_1^\dagger(x,\phi)|_{x=0,\phi\in(3\pi/2,2\pi)}$$

$$= x\partial_x \bar{P}_1^\dagger(r = r_1 + x\cos\phi, t = t_1 + x\sin\phi; r_1, t_1; r_1, 0)|_{x=0,\phi\in(3\pi/2,2\pi)}$$

$$= \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1}}_{<0} + x\sin\phi \times \underbrace{\partial_t \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=t_1}}_{=0} < 0. \tag{6.5.28}$$

We conclude that we have shown case $[i]$ of Theorem 6.5.2 by showing that inequality (6.5.21) applies for all possible directions.

We now proceed to showing case $[ii]$ of Theorem 6.5.2, and in particular that inequalities (6.5.16), (6.5.17) and (6.5.18) lead to a local ESS for the mimic in the sense of (6.5.11) in Definition 6.5.1. Much like in case $[i]$, we show this by showing that

$$\bar{\mathcal{P}}_2^\dagger(x,\phi) - \bar{\mathcal{P}}_2^\dagger(0,\phi) < 0 \text{ for all } \phi \in [0,\pi], \tag{6.5.29}$$

where $\bar{\mathcal{P}}_2^\dagger$ represents the average type-2 mutant fitness expressed in terms of polar coordinates x, ϕ. That is

$$\bar{\mathcal{P}}_2^\dagger : \mathbb{R}^{\geq 0} \times [0,\pi] \to \mathbb{R}^{\geq 0} : \ \bar{\mathcal{P}}_2^\dagger(x,\phi) := \bar{P}_2^\dagger(r = r_1 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_1, 0) \tag{6.5.30}$$

with

$$(x,\phi) \to (r,t) : \ \ r = r_1 + x\cos\phi \text{ and } t = x\cos\phi. \tag{6.5.31}$$

We show (6.5.29) by showing that it applies for cases $\phi = 0$, $\phi \in (0, \pi/2)$, $\phi = \pi/2$, $\phi \in (\pi/2, \pi)$ and $\phi = \pi$ individually.

If $\phi = 0$ mutation is along the r-direction so that

$$\bar{\mathcal{P}}_2^\dagger(x,\phi=0) - \bar{\mathcal{P}}_2^\dagger(0,\phi=0) = \bar{P}_2^\dagger(r,0;r_1,t_1;r_1,0) - \bar{P}_2^\dagger(r_1,0;r_1,t_1;r_1,0)$$

$$\approx \underbrace{(r - r_1)}_{>0} \times \underbrace{\vec{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{<0} < 0. \tag{6.5.32}$$

If $\phi \in (0, \pi/2)$ we have

$$\bar{\mathcal{P}}_2^\dagger(x,\phi) - \bar{\mathcal{P}}_2^\dagger(0,\phi) \approx x\partial_x \bar{\mathcal{P}}_2^\dagger(x,\phi)|_{x=0,\phi\in(0,\pi/2)}$$

$$= x\partial_x \bar{P}_2^\dagger(r = r_1 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_1, 0)|_{x=0,\phi\in(0,\pi/2)}$$

$$= \underbrace{x\cos\phi}_{>0} \times \underbrace{\vec{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{<0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\vec{\partial}_t \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{<0} < 0. \tag{6.5.33}$$

If $\phi = \pi/2$ mutation is solely along the t-direction so that

$$\bar{\mathcal{P}}_2^\dagger(x,\phi=\pi/2) - \bar{\mathcal{P}}_2^\dagger(0,\phi=\pi/2) = \bar{P}_2^\dagger(r_1,t;r_1,t_1;r_1,0) - \bar{P}_2^\dagger(r_1,0;r_1,t_1;r_1,0)$$

$$\approx \underbrace{(t - t_1)}_{>0} \times \underbrace{\vec{\partial}_t \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{<0} < 0. \tag{6.5.34}$$

If $\phi \in (\pi/2, \pi)$ the (average) incremental difference between mutant and mimic fitness is negative since

$$\bar{\mathcal{P}}_2^\dagger(x,\phi) - \bar{\mathcal{P}}_2^\dagger(0,\phi) \approx x\partial_x \bar{\mathcal{P}}_2^\dagger(x,\phi)|_{x=0,\phi\in(\pi/2,\pi)}$$

$$= x\partial_x \bar{P}_2^\dagger(r = r_1 + x\cos\phi, t = x\sin\phi; r_1, t_1; r_1, 0)|_{x=0,\phi\in(\pi/2,\pi)}$$

$$= \underbrace{x\cos\phi}_{<0} \times \underbrace{\overleftarrow{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{>0} + \underbrace{x\sin\phi}_{>0} \times \underbrace{\overrightarrow{\partial}_t \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{<0} < 0. \tag{6.5.35}$$

Finally, if $\phi = \pi$ mutation is solely along the r-direction so that $x = r - r_1 < 0$. We have

$$\begin{aligned}\bar{\mathcal{P}}_2^\dagger(x,\phi) - \bar{\mathcal{P}}_2^\dagger(0,\phi) =& \bar{P}_2^\dagger(r,0;r_1,t_1;r_1,0) - \bar{P}_2^\dagger(r_1,0;r_1,t_1;r_1,0)\\ \approx& \underbrace{(r-r_1)}_{<0} \times \underbrace{\overleftarrow{\partial}_r \bar{P}_2^\dagger(r,t;r_1,t_1;r_1,0)|_{r=r_1,t=0}}_{>0} < 0,\end{aligned} \tag{6.5.36}$$

which concludes our proof of (6.5.29) and of case (ii) of Theorem 6.5.2. □

For the remainder of the section we focus on making explicit the conditions for (local) ESS that are provided in Theorem 6.5.2 by substituting $\bar{P}_1^\dagger$ as given through (6.5.8) into (6.5.12), (6.5.13), (6.5.14), (6.5.15) and $\bar{P}_2^\dagger$ as given through (6.5.9) into (6.5.16), (6.5.17) and (6.5.18). We begin with (6.5.12). Evaluating $\partial_t \bar{P}_1^\dagger(r,t;r_1,t_1;r_1,0)$ at $(r,t) = (r_1,t_1)$ and scaling through by $(1-\bar{\gamma})F_1(t_1)/[D(r_1)K_1(t_1)]$ we arrive at equality

$$\left[\frac{F_1'(t_1)}{F_1(t_1)} - \frac{K_1'(t_1)}{K_1(t_1)}\right]\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)}d\gamma - a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+1}}{Q(I)}d\gamma = 0. \tag{6.5.37}$$

Strategies satisfying (6.5.37) must also satisfy (6.5.13), which on account of (6.5.8) reads

$$\begin{aligned}&\left[\frac{K_1''(t_1)}{K_1(t_1)} - \frac{F_1''(t_1)}{F_1(t_1)}\right]\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)}d\gamma\\ &+2a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{F_1'(t_1)}{F_1(t_1)}\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+1}d\gamma\\ &-2a^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\int_0^1 \frac{(Q'(I))^2}{Q^3(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+2}d\gamma\\ &+a^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\int_0^1 \frac{Q''(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+2}d\gamma\\ &+a\mathcal{I}_1^2\frac{H''(t_1)}{H(t_1)}\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+1}d\gamma > 0.\end{aligned} \tag{6.5.38}$$

Using the same scaling factor we can show that (6.5.14) amounts to

$$-\frac{D'(r_1)}{D(r_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)}d\gamma - \left[a\mathcal{I}_1\frac{D'(r_1)}{D(r_1)} - (1-a)\mathcal{I}_1 S'(0)\right]\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+1}d\gamma$$

$$+\mathcal{I}_2 S'(0)\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})}d\gamma > 0 \tag{6.5.39}$$

and similarly that (6.5.15) reads

$$-\frac{D'(r_1)}{D(r_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}}{Q(I)}d\gamma - \left[a\mathcal{I}_1\frac{D'(r_1)}{D(r_1)} + (1-a)\mathcal{I}_1 S'(0)\right]\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})+1}d\gamma$$

$$-\mathcal{I}_2 S'(0)\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})}d\gamma < 0. \tag{6.5.40}$$

Using the scaling factor $\bar{\gamma}F_2(0)/[D(r_1)K_2(0)]$ for the mimic it can be shown that (6.5.16) amounts to

$$\left[\frac{F_2'(0)}{F_2(0)} - \frac{K_2'(0)}{K_2(0)}\right]\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma$$

$$-a\mathcal{I}_2\frac{H'(0)}{H(0)}\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}+1}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma < 0. \tag{6.5.41}$$

Using the same scaling factor it is also straightforward to show that (6.5.17) and (6.5.18) amount to

$$-\frac{D'(r_1)}{D(r_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}}{Q(I)}d\gamma - \left[a\mathcal{I}_2\frac{D'(r_1)}{D(r_1)} - (1-a)\mathcal{I}_2 S'(0)\right]\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}+1}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma$$

$$+\mathcal{I}_1 S'(0)\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})}d\gamma > 0 \tag{6.5.42}$$

and

$$-\frac{D'(r_1)}{D(r_1)}\int_0^1 \frac{\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta}(1-\bar{\gamma}-1)}{Q(I)}d\gamma - \left[a\mathcal{I}_2\frac{D'(r_1)}{D(r_1)} + (1-a)\mathcal{I}_2 S'(0)\right]\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}+1}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma$$

$$-\mathcal{I}_1 S'(0)\int_0^1 \frac{Q'(I)}{Q^2(I)}\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})}d\gamma < 0. \tag{6.5.43}$$

It should be remarked that cases with non-zero rates of background mortality ($\lambda > 0$) are not considered in the context of Batesian mimicry so as to avoid excessive complexity. The reader is directed to chapter 4 for an account of this with prey populations consisting of a single species.

6.6 Eco-evolutionary stability and the confluent hypergeometric function

In this section we explore some of the interesting mathematical consequences that occur when considering an exponentially-decaying Q function along with the following biologically plausible functional forms relating

to predator generalisation

$$L(r) = \frac{1}{1+\exp(-r)}; \quad H(t) = t - t_c; \quad \lambda = 0$$

$$S(|r-r_1|) = \max\left(1 - v|r-r_1|, 0\right); \quad Q(I) = q_0 \exp(-qI). \tag{6.6.1}$$

We leave the remaining forms general for the time-being. The aversive information of (6.3.10) with (6.6.1) in place can be expressed as the difference of two terms, only one of which depends on γ. We have

$$I(\gamma) = (1-\gamma)\mathcal{I}_1 + \gamma\mathcal{I}_2 = \frac{N}{n}\frac{t_1 - t_c}{1+\exp(-r_1)} - \frac{N}{n}\frac{\gamma t_1}{1+\exp(-r_1)}, \tag{6.6.2}$$

where in the equal payoff condition (6.4.2) we can factor out the exponential that involves the first factor on the RHS to obtain the simplified expression

$$\frac{F_1(t_1)}{K_1(t_1)}\frac{1}{1-\bar{\gamma}}\int_0^1 \gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}\exp\left(-\frac{q\frac{N}{n}t_1\gamma}{1+\exp(-r_1)}\right)d\gamma$$

$$-\frac{F_2(0)}{K_2(0)}\frac{1}{\bar{\gamma}}\int_0^1 \gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}\exp\left(-\frac{q\frac{N}{n}t_1\gamma}{1+\exp(-r_1)}\right)d\gamma = 0. \tag{6.6.3}$$

We should remark that the integrals appearing on either side of the latter equality coincide with the integral representations of the family of *confluent hypergeometric functions*, which are denoted $M(x,y,z)$ in the standard handbook of mathematical functions Abramowitz and Stegun (1964). We have

$$M(x,y;z) = \frac{\Gamma(y)}{\Gamma(x)\Gamma(y-x)}\int_0^1 \exp(zu)u^{x-1}(1-u)^{y-x-1}. \tag{6.6.4}$$

Setting

$$x \leftrightarrow \delta\bar{\gamma}, \quad y \leftrightarrow \delta, \quad z \leftrightarrow -\frac{q\frac{N}{n}t_1}{1+\exp(-r_1)} \tag{6.6.5}$$

and using the relation $\Gamma(x+1) = x\Gamma(x)$ we observe that equality (6.6.3) amounts to

$$\frac{F_1(t_1)}{K_1(t_1)}M(\delta\bar{\gamma}, \delta+1; z) - \frac{F_2(0)}{K_2(0)}M(\delta\bar{\gamma}+1, \delta; z) = 0. \tag{6.6.6}$$

The transformed condition for equal payoffs given in (6.6.3) is consistent under small perturbations in the mimic and model proportions if it holds together with (6.4.7). We remark that the latter cannot be expressed in terms of the confluent hypergeometric function on account of the logarithms present in the

integrands of the last two terms. Through (6.6.1) condition (6.4.7) transforms to

$$\frac{\Gamma(\delta\bar{\gamma})\Gamma\left(\delta(1-\bar{\gamma})\right)}{\bar{\gamma}(1-\bar{\gamma})\Gamma(\delta)}\frac{F_2(0)}{K_2(0)}M(\delta\bar{\gamma}+1,\delta+1,z)$$

$$+\frac{\delta}{1-\bar{\gamma}}\frac{F_1(t_1)}{K_1(t_1)}\int_0^1\exp(z\gamma)\log\left(\frac{\gamma}{1-\gamma}\right)\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\gamma)}d\gamma$$

$$-\frac{\delta}{\bar{\gamma}}\frac{F_2(0)}{K_2(0)}\int_0^1\exp(z\gamma)\log\left(\frac{\gamma}{1-\gamma}\right)\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma>0. \tag{6.6.7}$$

The second equality necessary for achieving eco-evolutionary stability is provided by (6.5.37), which account of (6.6.1) reads

$$\left[\frac{F_1'(t_1)}{F_1(t_1)}-\frac{K_1'(t_1)}{K_1(t_1)}\right]\int_0^1\exp\left(-\frac{q\frac{N}{n}t_1\gamma}{1+\exp(-r_1)}\right)\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta\bar{\gamma}}d\gamma$$

$$-a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\int_0^1\exp\left(-\frac{q\frac{N}{n}t_1\gamma}{1+\exp(-r_1)}\right)\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta\bar{\gamma}+1}d\gamma=0. \tag{6.6.8}$$

Working as before it is straightforward to show that the latter can be expressed in terms of the hypergeometric function so that it reads

$$\left[\frac{F_1'(t_1)}{F_1(t_1)}-\frac{K_1'(t_1)}{K_1(t_1)}\right]M(\delta\bar{\gamma},\delta+1;z)-a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{\delta(1-\bar{\gamma})+1}{\delta+1}M(\delta\bar{\gamma},\delta+2;z)=0. \tag{6.6.9}$$

In fact all remaining stability conditions (excluding ecological stability) can be expressed in terms of hypergeometric functions. For the first type we have stability in the t-direction guaranteed when the equilibrium satisfies

$$\left[\frac{K_1''(t_1)}{K_1(t_1)}-\frac{F_1''(t_1)}{F_1(t_1)}\right]M(\delta\bar{\gamma},\delta+1,z)-2aq\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{F_1'(t_1)}{F_1(t_1)}\frac{\delta(1-\bar{\gamma})+1}{\delta+1}M(\delta\bar{\gamma},\delta+2,z)$$

$$-a^2q^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\frac{\left(\delta(1-\bar{\gamma})+2\right)\left(\delta(1-\bar{\gamma})+1\right)}{(\delta+2)(\delta+1)}M(\delta\bar{\gamma},\delta+3,z)>0. \tag{6.6.10}$$

Still for the first type, stability in the r-direction holds if (6.5.39) holds, which amounts to

$$-\frac{D'(r_1)}{D(r_1)}M(\delta\bar{\gamma},\delta+1,z)+\left[aq\mathcal{I}_1\frac{D'(r_1)}{D(r_1)}+(1-a)qv\mathcal{I}_1\right]\frac{\delta(1-\bar{\gamma})+1}{\delta+1}M(\delta\bar{\gamma},\delta+2,z)$$

$$-qv|\mathcal{I}_2|\frac{\delta\bar{\gamma}}{\delta+1}M(\delta\bar{\gamma}+1,\delta+2,z)>0 \tag{6.6.11}$$

and in addition, if (6.5.40) holds, which amounts to

$$-\frac{D'(r_1)}{D(r_1)}M(\delta\bar{\gamma},\delta+1,z)+\left[aq\mathcal{I}_1\frac{D'(r_1)}{D(r_1)}-(1-a)qv\mathcal{I}_1\right]\frac{\delta(1-\bar{\gamma})+1}{\delta+1}M(\delta\bar{\gamma},\delta+2,z)$$

$$+qv|\mathcal{I}_2|\frac{\delta\bar{\gamma}}{\delta+1}M(\delta\bar{\gamma}+1,\delta+2,z)<0. \tag{6.6.12}$$

For the second type stability against better-defended mutants is guaranteed through (6.5.41), which reads

$$\left[\frac{F_2'(0)}{F_2(0)} - \frac{K_2'(0)}{K_2(0)}\right] M(\delta\bar{\gamma}+1, \delta+1, z) + aq|\mathcal{I}_2|\frac{H'(0)}{H(0)}\frac{\delta\bar{\gamma}+1}{\delta+1}M(\delta\bar{\gamma}+2, \delta+2, z) < 0. \tag{6.6.13}$$

Along the r-direction for the second type we have (6.5.42) for non-invasion of the less conspicuous mutant type

$$-\frac{D'(r_1)}{D(r_1)}M(\delta\bar{\gamma}+1, \delta+1, z) - \left[aq|\mathcal{I}_2|\frac{D'(r_1)}{D(r_1)} + (1-a)qv|\mathcal{I}_2|\right]\frac{\delta\bar{\gamma}+1}{\delta+1}M(\delta\bar{\gamma}+2, \delta+2, z)$$

$$+qv\mathcal{I}_1\frac{\delta(1-\bar{\gamma})}{\delta+1}M(\delta\bar{\gamma}+1, \delta+2, z) > 0 \tag{6.6.14}$$

and (6.5.43) for non-invasion of the more conspicuous type

$$-\frac{D'(r_1)}{D(r_1)}M(\delta\bar{\gamma}+1, \delta+1, z) - \left[aq|\mathcal{I}_2|\frac{D'(r_1)}{D(r_1)} - (1-a)qv|\mathcal{I}_2|\right]\frac{\delta\bar{\gamma}+1}{\delta+1}M(\delta\bar{\gamma}+2, \delta+2, z)$$

$$-qv\mathcal{I}_1\frac{\delta(1-\bar{\gamma})}{\delta+1}M(\delta\bar{\gamma}+1, \delta+2, z) < 0. \tag{6.6.15}$$

6.7 An example of eco-evolutionarily stable Batesian mimicry

From the previous section it is clear that although eco-evolutionarily stable solutions are likely to manifest as pairs of curves in the $r_1, t_1, \bar{\gamma}$ - plane these are hard to evaluate explicitly. In this section we demonstrate that a pair of strategies for the model and mimic are eco-evolutionarily stable in the sense detailed thus far and proceed to demonstrating that these lie on an extended continuum. We begin by specifying the remaining functional forms that relate to prey behaviour in a non-stop style for ease of reference. We have

Conversions

$$\hat{M}(\delta\bar{\gamma}, \delta, z) = \int_0^1 \exp(\gamma z)\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma$$

$$z = -\frac{q\frac{N}{n}t_1}{1+\exp(-r_1)}$$

Ecological stability

$$\boxed{\frac{F_1(t_1)}{K_1(t_1)}\hat{M}(\delta\bar{\gamma}, \delta+1; z) - \frac{F_2(0)}{K_2(0)}\hat{M}(\delta\bar{\gamma}+1, \delta+1; z) = 0} \tag{6.7.1}$$

$$\frac{\Gamma(\delta\bar{\gamma})\Gamma\left(\delta(1-\bar{\gamma})\right)}{\bar{\gamma}(1-\bar{\gamma})\Gamma(\delta)}\frac{F_2(0)}{K_2(0)}\hat{M}(\delta\bar{\gamma}+1,\delta+1,z)$$

$$+\frac{\delta}{1-\bar{\gamma}}\frac{F_1(t_1)}{K_1(t_1)}\int_0^1\exp(z\gamma)\log\left(\frac{\gamma}{1-\gamma}\right)\gamma^{\delta\bar{\gamma}-1}(1-\gamma)^{\delta(1-\bar{\gamma})}d\gamma$$

$$-\frac{\delta}{\bar{\gamma}}\frac{F_2(0)}{K_2(0)}\int_0^1\exp(z\gamma)\log\left(\frac{\gamma}{1-\gamma}\right)\gamma^{\delta\bar{\gamma}}(1-\gamma)^{\delta(1-\bar{\gamma})-1}d\gamma>0 \tag{6.7.2}$$

Model ESS

$$\boxed{\left[\frac{F_1'(t_1)}{F_1(t_1)}-\frac{K_1'(t_1)}{K_1(t_1)}\right]\hat{M}(\delta\bar{\gamma},\delta+1;z)-a\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{\delta(1-\bar{\gamma})+1}{\delta+1}\hat{M}(\delta\bar{\gamma},\delta+2;z)=0.} \tag{6.7.3}$$

$$\left[\frac{K_1''(t_1)}{K_1(t_1)}-\frac{F_1''(t_1)}{F_1(t_1)}\right]\hat{M}(\delta\bar{\gamma},\delta+1,z)-2aq\mathcal{I}_1\frac{H'(t_1)}{H(t_1)}\frac{F_1'(t_1)}{F_1(t_1)}\frac{\delta(1-\bar{\gamma})+1}{\delta+1}\hat{M}(\delta\bar{\gamma},\delta+2,z)$$

$$-a^2q^2\mathcal{I}_1^2\left(\frac{H'(t_1)}{H(t_1)}\right)^2\frac{\left(\delta(1-\bar{\gamma})+2\right)\left(\delta(1-\bar{\gamma})+1\right)}{(\delta+2)(\delta+1)}\hat{M}(\delta\bar{\gamma},\delta+3,z)>0 \tag{6.7.4}$$

$$-\frac{D'(r_1)}{D(r_1)}\hat{M}(\delta\bar{\gamma},\delta+1,z)+\left[aq\mathcal{I}_1\frac{D'(r_1)}{D(r_1)}+(1-a)qv\mathcal{I}_1\right]\frac{\delta(1-\bar{\gamma})+1}{\delta+1}\hat{M}(\delta\bar{\gamma},\delta+2,z)$$

$$-qv|\mathcal{I}_2|\frac{\delta\bar{\gamma}}{\delta+1}\hat{M}(\delta\bar{\gamma}+1,\delta+2,z)>0 \tag{6.7.5}$$

$$-\frac{D'(r_1)}{D(r_1)}\hat{M}(\delta\bar{\gamma},\delta+1,z)+\left[aq\mathcal{I}_1\frac{D'(r_1)}{D(r_1)}-(1-a)qv\mathcal{I}_1\right]\frac{\delta(1-\bar{\gamma})+1}{\delta+1}\hat{M}(\delta\bar{\gamma},\delta+2,z)$$

$$+qv|\mathcal{I}_2|\frac{\delta\bar{\gamma}}{\delta+1}\hat{M}(\delta\bar{\gamma}+1,\delta+2,z)<0 \tag{6.7.6}$$

Mimic ESS

$$\left[\frac{F_2'(0)}{F_2(0)}-\frac{K_2'(0)}{K_2(0)}\right]\hat{M}(\delta\bar{\gamma}+1,\delta+1,z)+aq|\mathcal{I}_2|\frac{H'(0)}{H(0)}\frac{\delta\bar{\gamma}+1}{\delta+1}\hat{M}(\delta\bar{\gamma}+2,\delta+2,z)<0 \tag{6.7.7}$$

$$-\frac{D'(r_1)}{D(r_1)}\hat{M}(\delta\bar{\gamma}+1,\delta+1,z)-\left[aq|\mathcal{I}_2|\frac{D'(r_1)}{D(r_1)}+(1-a)qv|\mathcal{I}_2|\right]\frac{\delta\bar{\gamma}+1}{\delta+1}\hat{M}(\delta\bar{\gamma}+2,\delta+2,z)$$

$$+qv\mathcal{I}_1\frac{\delta(1-\bar{\gamma})}{\delta+1}\hat{M}(\delta\bar{\gamma}+1,\delta+2,z)>0 \tag{6.7.8}$$

$$-\frac{D'(r_1)}{D(r_1)}\hat{M}(\delta\bar{\gamma}+1,\delta+1,z)-\left[aq|\mathcal{I}_2|\frac{D'(r_1)}{D(r_1)}-(1-a)qv|\mathcal{I}_2|\right]\frac{\delta\bar{\gamma}+1}{\delta+1}\hat{M}(\delta\bar{\gamma}+2,\delta+2,z)$$

$$-qv\mathcal{I}_1\frac{\delta(1-\bar{\gamma})}{\delta+1}\hat{M}(\delta\bar{\gamma}+1,\delta+2,z)<0. \tag{6.7.9}$$

Before proceeding to a worked example of these conditions (see Figure 6.3) it is of interest to make two remarks. The first is that the conditions for evolutionary stability for the model and mimic have a certain physical significance, which becomes apparent when regarding the individual terms on the LHSs as marginal differences in fitness between the mutant and the model/mimic. For instance, looking at conditions for non-invasion along r e.g. (6.7.5), (6.7.6) for the model and (6.7.7), (6.7.9) for the mimic we revise our interpretation as follows. In (6.7.5) we deduce (subject to scaling by $-\delta r$) that the first term (+) represents a marginal fitness advantage of the mutant compared to the model associated with being less detectable; the second term (−) represents a marginal advantage associated with being less memorable as an aversive type; the third term (−) represents a fitness disadvantage associated with the mutant type looking marginally unlike the model, whose appearance is registered by the predator as aversive and finally the fourth term (+) represents the advantage associated with the mutant type looking unlike an attractive mimic type. We examined the terms of (6.7.5) merely as a demonstration; the reader is encouraged to scrutinise the remaining ESS conditions in this manner.

Although fundamentally this term-by-term marginal fitness interpretation is much the like in the single species case (see chapter 2) there are two distinguishing differences. The first is that there are four terms making up the LHSs of (6.7.5), which is attributed to the fact that the prey population is now made up of two types with common appearance (perfect mimicry) and therefore a slightly less/more conspicuous mutant pays a price for not looking like the aversive models and gains an advantage for looking unlike the attractive mimics. Perhaps the reader may find it more intuitive to view these two terms as a single term by looking at the perceived aversiveness of the mimicry complex as a whole and the associated advantage/disadvantage of the mutant as derived from that.

We proceed by assigning the same functional forms for the model and the mimic and set

$$F_1(t)=f_{01}\exp(-f_1t);\quad F_2(t)=f_{02}\exp(-f_2t)$$

$$K_1(t)=\frac{k_{01}}{1+k_1t_1};\quad K_2(t)=\frac{k_{02}}{1+k_2t}\quad D(r)=\frac{1}{1+\exp(-r)}. \tag{6.7.10}$$

A stable point solution

We proceed by setting $a=0$, which simplifies the conditions for eco-evolutionary stability conditions (6.7.1) through to (6.7.9) considerably. For this example we also wish to fully specify the distribution of mimics in the habitat. We set $\alpha=2$ and $\beta=5$ which implies that $\delta=7$ and furthermore that the average proportion of mimics is $\bar{\gamma}=2/7$. In chapter 9 of Ruxton et al. (2019) it is mentioned that real Batesian mimicry systems are likelier to be stable (stability is used in a more general sense) when the models are considerably larger in density and their aversiveness is also considerable compared with the mimics. We account for the first requirement by considering a distribution from the Beta family that gives $\bar{\gamma}=2/7$ and $1-\bar{\gamma}=5/7$. Since we take it that mimics are completely undefended making the latter requirement superfluous; we will

still consider cases where the model toxicity is safely beyond the critical level t_c, which we arbitrarily set to $t_c = 0.25$.

We impose that $q\frac{N}{n}t_1 = 1$ and set $r_1 = 1$ so that

$$z = \frac{1}{1+\exp(-1)} \tag{6.7.11}$$

so that now it is straightforward to establish that when $a = 0$ the level of defence satisfying (6.7.3) is fixed for all levels of conspicuousness. This is given as

$$\left[-f_1 + \frac{k_1}{1+k_1t_1}\right] M(2,8,z) = 0 \;\; \Leftrightarrow t_1 = \frac{1}{f_1} - \frac{1}{k_1}. \tag{6.7.12}$$

We should remark that checking (6.7.4) for the model is superfluous since this condition is automatically satisfied at t_1 given through (6.7.12). Indeed, the LHS of (6.7.4) with $a = 0$ reads

$$\left[-f_1^2 + \frac{2k_1^2}{(1+k_1t_1)^2}\right] \hat{M}(2,8,z). \tag{6.7.13}$$

The equal payoffs condition (6.7.1) with $k_{01} = k_{02}$ - the latter suggests that models and mimics are equally likely to escape capture when undefended - amounts to

$$f_{01}\exp(-f_1t_1)(1+k_1t_1)\hat{M}(2,8,z) = f_{02}\hat{M}(3,8,z) \tag{6.7.14}$$

and setting $k_1 = 1$ gives

$$f_1 = \frac{1}{e}\frac{f_{01}}{f_{02}}\frac{\hat{M}(2,8,z)}{\hat{M}(3,8,z)}\exp(f_1). \tag{6.7.15}$$

Solutions to equalities involving a certain variable and the same variable as an exponent can be solved by means of the following ansatz

$$x = a + b\exp(cx) \quad \Rightarrow x = a - \frac{1}{c}W(-bc\exp(ac)) \tag{6.7.16}$$

with a, b and c generally being complex constants and W being of any integer order. The ansatz implies

$$f_1 = -W\left(-\frac{1}{e}\frac{f_{01}}{f_{02}}\frac{\hat{M}(2,8,z)}{\hat{M}(3,8,z)}\right) \tag{6.7.17}$$

and solutions are sensible (real and positive) providing

$$0 < \frac{f_{01}}{f_{02}}\frac{\hat{M}(2,8,z)}{\hat{M}(3,8,z)} < 1.$$

Since

$$\frac{\hat{M}(2,8,z)}{\hat{M}(3,8,z)} = \frac{\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma(1-\gamma)^5 d\gamma}{\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^4 d\gamma} \approx 2.7353 \tag{6.7.18}$$

setting $f_{01}/f_{02} = 1/6$ achieves the desired result and also suggests that the model fecundity (in absence of investment in secondary defences) is smaller than the mimic fecundity, which is a safe assumption to make.

From (6.7.17) we have

$$f_1 \approx -W\left(\frac{1}{e}\frac{2.7353}{6}\right) \approx 0.2061 \tag{6.7.19}$$

and from (6.7.12) we deduce that the model toxicity is

$$t_1 \approx \frac{1}{0.2061} - 1 \approx 3.8522, \tag{6.7.20}$$

which is aversive (and notably so compared with the critical level at 0.25). For consistency we should remark that the recovered values are consistent with the equal payoffs condition in (6.7.14), which reads

$$\frac{f_{01}}{f_{02}}\frac{\hat{M}(2,8,z)}{\hat{M}(3,8,z)}\exp(-f_1 t_1)(1+k_1 t_1) \approx \frac{2.7353}{6}\exp\left(-0.2061\times 3.8522\right)(1+3.8522) \approx 1.0000, \tag{6.7.21}$$

as required.

With the parameter values specified thus far the equal payoffs condition (6.7.14) is consistent under small perturbations in the mimic/model proportions if (6.7.2) holds. Th latter reads

$$+\frac{49}{300}\Gamma(2)\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^4 d\gamma$$

$$\frac{49}{30}\exp(-0.2061\times 3.8522)\times(1+3.8522)\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma(1-\gamma)^5\log\left(\frac{\gamma}{1-\gamma}\right)d\gamma$$

$$-\frac{49}{2}\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^4\log\left(\frac{\gamma}{1-\gamma}\right)d\gamma \approx 0.02344 > 0. \tag{6.7.22}$$

We have therefore established that the pair of strategies $(r_1, t_1) \approx (1, 3.8522)$ for the model and $(r_1, 0) = (1, 0)$ for the mimic played in average proportions 5/7 and 2/7 respectively are ecologically stable in the sense of receiving equal payoffs $\bar{P}_1$ and $\bar{P}_2$ - condition (6.7.1) - in such a way that this condition is consistent under small perturbations in the mimic/model proportions -condition (6.7.2). In addition, we have also shown that model and mimic can resist invasion along t in the sense of (6.7.3), (6.7.4) and (6.7.7). We now proceed to identifying a suitable set of parameter values that are consistent with (6.7.5), (6.7.6) for the model and (6.7.8), (6.7.9) for the mimic.

Scaling (6.7.5) and (6.7.6) by $(1+\exp(-1))$ and setting $v = 8$ gives

$$-\exp(-1)\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma(1-\gamma)^5 d\gamma$$

$$+2.6\times 1\times\left(1-\frac{0.25}{3.8522}\right)\times\frac{3}{4}\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma(1-\gamma)^6 d\gamma$$

$$-2.6\times 1\times\frac{0.25}{3.8522}\times\frac{1}{4}\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^5 d\gamma$$

$$\approx -0.007336 + 0.02780 - 0.0001979 \approx 0.02027 > 0$$

as required. Doing the same for condition (6.7.6) gives

$$-0.007336 + 0.02780 + 0.0001979 \approx -0.03494 < 0. \tag{6.7.23}$$

It therefore remains for us to show that conditions (6.7.8) and (6.7.9) hold for the mimic. We have seen that with $a = 0$ the mimic satisfies (6.7.7) provided that $f_2 > k_2$. Consistent with this (6.7.7) amounts to

$$\begin{aligned}
&-\exp(-1)\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^4 d\gamma \\
&-2.6 \times \frac{0.25}{3.8522} \times \frac{3}{8}\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^3(1-\gamma)^4 d\gamma \\
&2.6 \times \left(1 - \frac{0.25}{3.8522}\right) \times \frac{5}{8}\int_0^1 \exp\left(\frac{-\gamma}{1+\exp(-1)}\right)\gamma^2(1-\gamma)^5 \\
&\approx -0.002682 - 0.0001644 + 0.007130 \approx 0.004284 > 0
\end{aligned} \tag{6.7.24}$$

Similarly, condition (6.6.15) is consistent with the above parameter values and amounts to

$$-0.002682 + 0.0001644 - 0.007130 \approx -0.009649 < 0.$$

We have demonstrated how one can generate examples of mimicry complexes that are consistent with the conditions for eco-evolutionary stability outlined in (6.7.1) through to (6.7.9) by choosing parameter values appropriately. As far as the functional forms of (6.6.1) and (6.7.10) are concerned, their efficacy has been discussed more carefully in Broom et al. (2006) as well as in Broom et al. (2008) and indeed elsewhere in this manuscript. In following with the presentation of the previous chapter (5) on Batesian mimicry and for the purpose of maintaining simplicity, we have imposed that models and mimics are fundamentally described through a common set of *functional forms* but allow for the specific *functions* to be different by assigning different parameter values - see (6.7.10) K_1 and K_2 have the same form but differ in terms of the parameters k_{01}, k_{02} and the k_1, k_2.

It should be clear to the reader that the example presented here is not designed as one of a kind (non-generic). First, we should remark that the process detailed above is general and can be made to work for a variety of instances. Indeed, setting $k_1 = 1$ was an arbitrary choice and the reader is invited to show that working examples can be generated using other values. The same applies to implemented restrictions of the type $k_{01} = k_{02}$ or specifically setting $f_{01}/f_{02} = 1/6$ (smaller values for this fraction are also viable options). In addition, our conclusions regarding the eco-evolutionary stability hold for arbitrary choices of the parameters f_2 and k_2 that satisfy the inequality $f_2 > k_2$ used for generating examples. Second, as we demonstrate in the supplementary analysis that follows the point solution we have uncovered in this section lies on a continuum of strategies (i.e. choices for r_1, t_1 and $\bar{\gamma}$ are drawn from intervals of values) that is eco-evolutionarily stable (see Figure 6.3). Specifically, we show that for a given choice of the parameters - here we have set $a = 0, f_1 \approx 0.2061, k_1 = 1, t_c = 0.25, qN/n \approx 0.2596, k_{01} = k_{02}, f_{01}/f_{02} = 1/6, v = 2.6, \delta = 7$ - there is a range of values that $(r_1, t_1, \bar{\gamma})$ can assume in the vicinity of $(r_1, t_1, \bar{\gamma}) = (1, 3.8522..., 2/7)$.

From point solutions to continua

In this section we demonstrate that for the specified parameter values the point solution discussed above can be shown to be eco-evolutionarily stable over a continuum for strategies r_1, t_1 and $\bar{\gamma}$. In particular that for $a = 0, f_1 \approx 0.2061, k_1 = 1, t_c = 0.25, qN/n \approx 0.2596, k_{01} = k_{02}, f_{01}/f_{02} = 1/6, v = 2.6$ and $\delta = 7$ the point solution $(r_1, t_1, \bar{\gamma}) = (1, 3.852..., 2/7)$ is eco-evolutionarily stable within the sub-region $[0.1145, 9.0000] \times 3.8522..., [1.965/7, 2.0466/7]$ of the strategy space.

We begin by remarking that condition (6.7.18) can be solved for values of r_1 and $\bar{\gamma}$ that lie on the curve shown in Figure 6.2, which on account of a being zero is contained on the $t_1 \approx 3.8522...$ cross section - indeed, it should be clear from condition (6.7.3) for the model that setting $a = 0$ results in curves that are planar in $\mathbb{R}^3$. The curve shown in Figure 6.2 is defined implicitly through condition (6.7.12) for the model, which amounts to $t_1 = 1/f_1 - 1$ with f_1 defined through the ecological equilibrium condition (6.7.1), which in turn amounts to (6.7.18) and (6.7.12). The form of the continuum in Figure 6.2 was determined numerically by varying $\bar{\gamma}$ in the vicinity of $\bar{\gamma} = 2/7$ and then identifying the value of r_1 such that equality (6.7.18) is maintained. The analytic expression for the extended equilibrium (ecological and evolutionary) of Figure 6.3 is given implicitly through the aforementioned equalities and is not simplified further.

Now that we have established that the points on the curve of Figure 6.2 satisfy the ecological and evolutionary equilibrium conditions it remains for us to establish which sections of the curve satisfy which of the remaining stability conditions from (6.6) through to (6.7.9). Stability in the t-direction is not influenced by changes in the strategies r_1 and $\bar{\gamma}$ and therefore the associated conditions for the model (6.7.4) and mimic (6.7.7) hold true on all points along the curve. Table 6.1 shows a number of coordinate points that are contained in the curve shown in Figure 6.3. The red X-marks and green check-marks indicate whether the coordinates satisfy the conditions for eco-evolutionary stability. From Table 6.1 it is clear that although for those lower values of the conspicuousness the mimics risk invasion from the less conspicuous mutant types this seizes to be the case as the conspicuousness (and associated proportion of mimics) increases. While it has not been possible to show analytically that every point with $r_1 > 0.145$ along the continuum of Figure 6.3 is eco-evolutionarily stable we have done so for the sequence of points that are mentioned in Table 6.1. Furthermore, on account of the functional forms of (6.6.1) and (6.7.10) being smooth and the size of the perturbation around $\bar{\gamma} = 2/7$ is small (within 2% of the initial value) the reader may safely convince themselves by looking at the variation in the associated values of Table 6.1 that sections with $r_1 < 0.1145$ are unstable and that sections above that are stable.

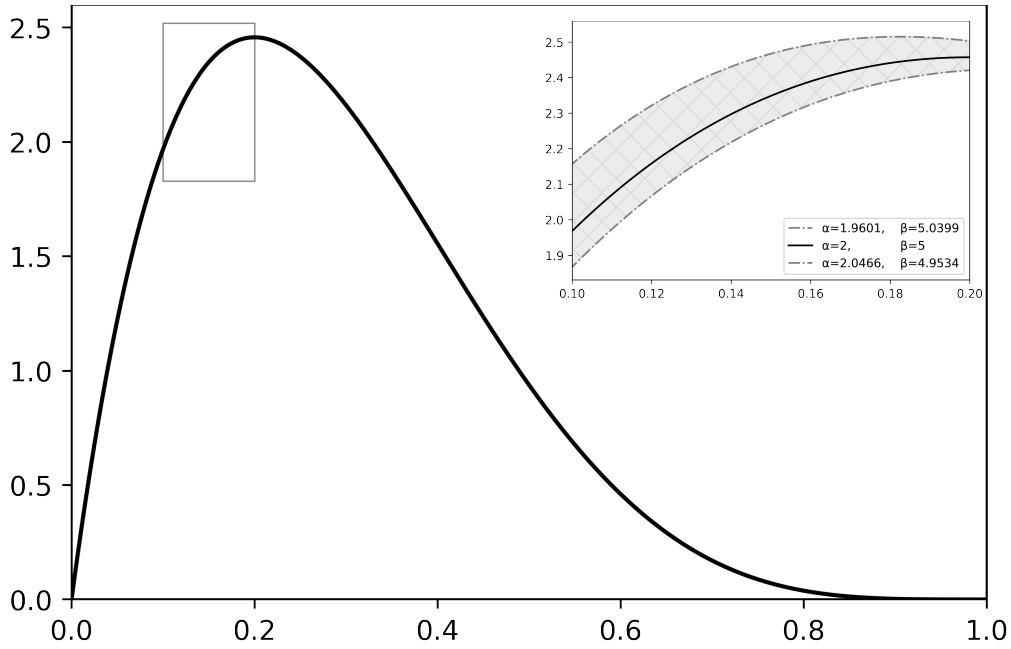

Figure 6.2: Solid black curve shows plot of probability density for the proportion of mimics (γ) in the population of prey, which is beta distributed with mean $\bar{\gamma} = 2/7$ and $\delta = 7$. The interior of the grey rectangle is shown in greater detail in the close-up figure on the top right, in which the shaded region represents a perturbation of the density function about the mean value $\bar{\gamma} = 2/7$ within the interval $(1.965/7, 2.045/7)$ keeping $\delta = 7$ fixed.

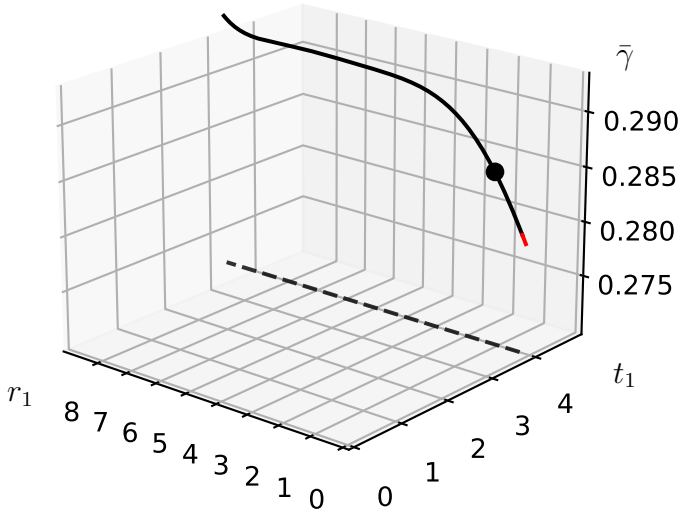

Figure 6.3: Solid black curve shows the continuum of eco-evolutionarily stable solutions that are defined in the vicinity of the initial example $(r_1, t_1, \bar{\gamma}) = (1, 3.8522..., 2/7)$, which is shown as a solid black marker. The red section of the curve shows that the continuum is unstable (mimic invaded by the less conspicuous type) for smaller values of the conspicuousness and becomes stable for increasing values (black section of the curve). The figure is to be viewed in tandem with Table 6.1 in which the exact values at each coordinate of the curve provide an indication of how stable each is and with respect to which potential invader.

The solid curve in Figure 6.3 consists of strategies satisfying both (6.7.1) and (6.7.3) as a continuum that is contained in the $t_1 \approx 3.8522$ plane. It is clear from the plot that the mean mimic proportion and associated conspicuousness are positively related along this curve. The confinement of the curve in the $t_1 \approx 3.8522$ plane is justified by our choosing $a = 0$ in the particular example. That is, while (6.7.3) in general imposes a restriction on all three variables r_1, t_1 and $\bar{\gamma}$ when setting $a = 0$ the last term on the LHS of the named equality is eliminated as is any restriction on r_1 and $\bar{\gamma}$ leading to a restriction solely on the variable t_1; the associated plane of values therefore represents those strategies that satisfy (6.7.3). The eco-evolutionarily stable curve is generated through the intersection of this plane with the surface of strategies (not plotted) that satisfy equality (6.7.1). To that end it is important to emphasize that by construction the eco-evolutionarily stable curve is defined within a fixed pair of endpoints associated with when the surface "enters" and when the surface "exits" the plane. These are located precisely at $(0.009, 1.9601/7, 3.8522)$ and $(9, 2.0466/7, 3.8522)$ so that the curve is not defined outside these bounds.

The relationship between $\bar{\gamma}$ and conspicuousness along this curve is increasing. We should recall that since $\bar{\gamma}$ represents the proportion of mimics a marginal increase in this quantity is paired with an overall decrease in the perceived aversiveness of the complex. It is therefore plausible that the models need to advertise their aversiveness more strongly (aversive learning is crucial for the maintenance of aposematism - see Mappes et al., 2005 but also the discussion below - and stronger signals could facilitate this). Interestingly, the continuum exhibits a plateau where the surging effect on perceived aversiveness (and survival) of a marginally increasing proportion of mimics cannot be counterbalanced by ever-increasing signal intensity and the solution ceases to be defined beyond a certain point - the level of defence satisfying (6.7.3) would have to increase to maintain ecological equilibrium (6.7.1) but this would violate (6.7.3). The more specific reason for the observed plateau is likely the assumed plateau in rates of prey detection ($D(r)$) and predator recollection ($L(r) = D(r)$); while this assumption is reasonable for $D(r)$ the plausibility of perfect predator recollection is more questionable. In fact, it would be interesting to consider worked examples of the above in which predator recollection is given by the product of the detection rate and a function that increases with prey conspicuousness.

r_1	$\delta\bar{\gamma}$	Ecol. stab	Model r-stab [L]	Model r-stab [R]	Mimic r-stab [L]	Mimic r-stab [R]	Stability
0.0009	1.9601	0.02865	0.008794	-0.05324	-0.0003000	-0.01595	X
0.0101	1.9605	0.02859	0.1292	-0.1731	-0.0002259	-0.01586	X
0.1145	1.965	0.02800	0.01103	-0.05017	0.0005582	-0.01487	✓
1	2.000	0.1013	0.02027	-0.03494	0.004284	-0.009649	✓
2.245	2.03	0.02001	0.02335	-0.02724	0.005673	-0.007096	✓
4.665	2.045	0.01854	0.02405	-0.02438	0.006052	-0.006173	✓
9.000	2.0466	0.09278	0.02410	-0.02410	0.006083	-0.006085	✓

Table 6.1: In any given row are shown the equilibrium conspicuousness and associated mean proportion of mimics for the worked example (i.e. $a = 0, f_1 \approx 0.2061, k_1 = 1, t_c = 0.25, qN/n \approx 0.2596, k_{01} = k_{02}, f_{01}/f_{02} = 1/6, v = 2.6$ and $\delta = 7$). Note that all such coordinates are contained in the curve of Figure 6.3. The remaining columns correspond to the quantities on LHSs of the conditions for bi-stability (6.7.1) through to (6.7.10) that are not trivially true (for instance, t-stability for the model (6.7.4) and mimic (6.7.7) are omitted). Red crosses in the far-right column indicate that the data point(s) in question violates at least one of the conditions for bi-stability - in our case the mimic types corresponding to the first two coordinates are invaded by the less conspicuous mutant types (6.7.8), while all other coordinates are stable, as indicated by the green check-marks.

6.8 Items for discussion

In the last couple of sections we have used the theory developed earlier in the chapter to demonstrate that models and mimics may co-exist in a sense that is stable both on the short-term ecological time-scale and on the longer evolutionary time-scale. This was achieved by considering a point in the (expanded) strategy space $(r_1, t_1, \bar{\gamma})$ and showing that it satisfies all the necessary conditions. The next step was to show that by marginally varying the assumed mean proportion of mimics $\bar{\gamma} = 2/7$ one can uncover a range of associated levels for the conspicuousness that would achieve the same result - see Table 6.1 and the continuum of Figure 6.3. While these results have been reached successfully it is worth noting a number of facts that bare significance not only to the calculations but also to observed complexes of Batesian mimicry. In particular, we contend that the stability of a Batesian mimicry complex is sensitive to the resident toxicity being sufficiently high (in this case it was greater than fifteen times the critical level) and to the average proportion of mimics being not too high (we have used a mean values of $\bar{\gamma} = 2/7$). We discuss our findings in the broader context of Batesian mimicry below.

Even in its original presentation (e.g. involving a single species - see chapter 2) the model of Broom et al. (2006) had been too complex for all of its parameters to be tested empirically and we expect this effect to be compounded in its extension to mimicry systems. While our success in predicting stable outcomes for Batesian mimicry systems is notable, there are a number of criticisms (and points for improvement) to be made and on different levels. For instance, in these last two chapters we have treated Batesian mimicry as a complex involving only two species, but there is ample evidence to suggest that Batesian mimics form *mimicry rings* in their natural habitats. In Ruxton et al. (2019) it is argued that perhaps a whole community approach, such as that taken in Pekár et al. (2017) is the more appropriate evolutionary context to study mimicry and we encourage the reader to consult this.

Furthermore, the laboratory experiments on avian predators (such are plausible candidates for the overall setup of Broom et al., 2006) conducted by Gamberale-Stille and Guilford (2004) and Skelhorn and Rowe (2006) seem to provide evidence in favour of the possibility that predators are able to selectively distinguish between noxious models and more palatable mimics after sampling. This sample-rejection behaviour could have repercussions (unknown precisely how large) on the stability of Batesian mimicry and is therefore worth considering. The works of Holen (2013) conclude that if a mimicry complex is unprofitable (such as in the example of this section) and alternative prey are available that predators are less likely to attack (stable), while if the system is profitable a taste-rejecting predator could reject models and consume more mimics (unstable). Interestingly, even though our model does not account for taste-rejection, these conclusions are not misaligned with the reasoning in our last example in which we predict stability subject to the assumption that the models on average constitute a severely ($t_1 \approx 3.8522$ compared with the base-level of $t_c = 0.25$) toxic majority ($\bar{\gamma} \approx 2/7$).

Laboratory studies dating back to the start of the 20^{th} century have examined the genetics of Batesian mimicry with those of Fryer (1914), Punnett (1915) and Gabritschevsky (1924) providing invaluable precursory insight into the genetics of polymorphism (a common and important feature in many systems of mimicry encountered in the natural world). Since then the more recent employment of advanced genomics has lead to a number of significant breakthroughs and the reader is referred to Deshmukh et al. (2018) for a review of these. The mentioned studies and indeed the majority of those focusing on the genetics of Batesian mimicry have considered lepidopteran groups (including butterflies and moths); there is evidence to suggest that frogs (see Darst and Cummings, 2006 or Twomey et al., 2013) and/or snakes (see Kikuchi and Pfennig,

2010 and Kikuchi and Pfennig, 2012) to which the model of Broom et al. (2006) is more adept also form Batesian mimicry structures and more work would be required to understand the genetics in these taxa. It is worth mentioning the three extensive works by Charlesworth and Charlesworth of 1975 , which incorporated a large number of different elements into an early computer simulation model of these genetics. Indeed, it is conceivable that the effort to fully understand the genetics of mimicry is largely interdisciplinary (involving realms of developmental and molecular biology) and such a combined endeavour is still at the outset.

While the original model of Broom et al. (2006) had considered two distinct traits to describe the visual appearance of aposematic prey (conspicuousness and colouration), we have only considered the first. Although it is probably true that without this omission the extension of the model to prey populations consisting of two species would not have been possible it also becomes clear that an obvious area to further develop would be its inclusion. Interestingly, this point brings us to the discussion about multi versus single-trait mimicry. A number of questions remain unanswered regarding how predators decide to attack or not to attack prey animals and especially in instances where the rules that these employ involve the interactions of several visual traits. In our discussion we have assumed that the conspicuousness trait is also the (single) trait involved in deceiving the predator, but the reality may well be that there are other correlated traits that could influence this (including colouration and/or size). As mentioned earlier in the section, it is not fully understood whether more conspicuous signals are better recollected by predators and the assumption of perfect predator recollection could be revised to consider rates of predator recollection is that is not simply a scalar of the rate of detection, but is instead scaled by a function that increases with prey conspicuousness.

In closing, we focus the discussion on studies which have relied on mathematical models to examine Batesian mimicry. A detailed investigation would reveal that there is a limited number of such studies and that none deal with the evolutionary stability of mimicry in the sense of a local ESS. Nonetheless, the ones we discuss below (apologies for undue omissions) are noteworthy and provide insight into the aspects (particularly the dynamics) that are lacking in our own. After all, our own approach is static and the reader can doubtlessly benefit from supplementing their reading. The mathematical model of Kato and Takada (2019) takes a dynamical-systems theory approach to consider the model-mimic community dynamics of the prey population. Through bifurcation methods they numerically predict regions in the parameter space where ratio-dependent equilibria between models and mimics are to be expected. We should remark that while mutation is not present their notion of stability (which is explored to considerable depth) is limited to the dynamics and in a sense this approach is complementary to our own. The earlier work by Oaten et al. (1975) uses a signal detection theory approach to considers a joint population of models and mimics. Interestingly, this study can account for predators utilising sophisticated discriminative rules by regarding prey signals as k-dimensional vectors with normally-distributed traits of appearance. There are several works in the realm of signal detection theory but their inclusion is beyond our scope. Matessi and Cori (1972) develop a detailed mathematical model, which investigates the effect of selective predation on the frequency of a certain gene that induces Batesian mimicry. In this, a predator's decision to attack mimics depends on their relative abundance to models in the population and the authors predict a number of different outcomes, including both stable and unstable equilibria (as we also do), where in the latter case they predict a series of stable oscillations in the gene frequency about the equilibrium. Matessi later co-authored Matessi and Gimelfarb (2006), which explores the evolutionary stability of polymorphism in a continuous trait. The model is not about Batesian mimicry; although colour polymorphisms do arise in Batesian mimicry systems.

The more recent mathematical model-based study by Kikuchi et al. (2022) explores the effect on the stability of Batesian mimicry systems of specific predator population dynamics. It is perhaps the only

attempt to integrate community ecology and evolutionary ecology approaches to Batesian mimicry and is therefore worth noting. They make two predictions, which we also confirm here. (i) They predict that when the proportion of mimics was highest the number of attacks experienced by the complex as a whole increased and they cite the empirical works by Kristiansen et al. (2018) done on butterflies and Pfennig et al. (2007) to validate this. Indeed, this assumption is fundamentally built into (6.3.10), which determines the predator's decision on whether to mount an attack (through the perceived aversiveness of the complex). (ii) The more protected the models were the larger the predicted *mimic load* (proportion of mimics). As we have previously mentioned, the eco-evolutionarily stable proportion of mimics can increase along the continuum of Figure 6.3 providing the model conspicuousness was sufficiently high.

Much of the cited works discuss Batesian (and not Browerian) mimicry. This is mostly because Batesian mimicry is the main focus of this chapter but also because far fewer studies have been done specifically on automimics. Nonetheless, as reported in Ruxton et al., (2019) the extent to which automimics can invade a population of automodels "*depends on critically on the response of predators. If predators attack rates increase as cheats increase in frequency, then the mortality risk to cheats increases more quickly than it would for non-cheats because cheats are less likely to survive an attack*". The mathematical model by Svennungsen and Holen (2007) builds on that by Broom et al. (2005) and investigates the scenaria in which the shapes of the fecundity (e.g. log-concave, log-linear and log-convex) and survival functions (e.g. concave, convex and sigmoid) influence the possibility of an evolutionarily stable dimorphism. Their approach is insightful within itself and could be of additional interest to the empiricist due to its organised cataloguing of outcomes.

Closing remarks

In this thesis we discuss how prey individuals gain selective advantage by signalling their unprofitability to potential predators. This phenomenon, which had originally been described by Wallace as a process of "warning colouration" is presently referred to using the more general term "aposematism" and it is remarkable how our understanding of the process has advanced since its first observation. To that, we point out that the biology of defence is vast with different types of defence being present at different times relative to encounter and bearing an array of different benefits/costs to organisms that acquire them. The central game of this thesis is more adept to (but not restricted to) the modelling chemically-defended prey whose unpalatable toxins are transmitted to predators during or after encounter and whose presence is signalled through the external signalling cues (such as bright skin pigmentation). Upon first inspection the evolution of such a mechanism seems curious as it seems to provide no direct benefit to the prey individuals that acquire it. We model the process using a static approach in such a way that the values of their traits (signalling and defence) are realised as continuously quantifiable quantities that can be varied independently from one another and which together define a two-dimensional strategy space. Within this the aposematic behaviour of any one organism can be represented by a single point. It is also worth observing that while the central game of this thesis is non-linear and perhaps more multifaceted than the War of Attrition described in the first chapter, evolutionarily stable outcomes are manifest as a continuum in the strategy space.

For prey populations consisting of a single type we successfully demonstrate that evolutionarily stable levels of signal strength may exhibit both a positive or a negative relationship with the defence and further, that for a given level of signal strength there can be more than one optimal level of defence. As we discuss, empirical and model-based studies are conflicting regarding how aposematic traits are and should be related to one another in nature (this includes Broom et al., 2006), which renders the above results all the more relevant. Although it is true that the majority of works allude to a positive relationship between signalling and defence, this is by no means a definitive conclusion. Furthermore, even if this were the case, there is no one accepted mechanism by which this is facilitated (although a number of proposed mechanisms are considered). It is especially intriguing that using an exclusively static approach we can accommodate for such a vast range of outcomes.

By means of example we also explore the relationship between evolutionarily stable levels of defence and signal strength under various regimes of background mortality and colony size; notably all previous efforts have assumed predation to be the only source of death. These predictions are compared with novel simulations on populations of prey that are finite and subject to random local mutation in the context of a genetic algorithm model, which we define. For the first time the roles of absolute resident fitness, marginal mutant fitness and stochasticity are considered jointly in the evolution of prey traits and in so doing we discuss the importance of population size in the above. We contend that we have extended the scope of the celebrated model by Broom et al. (2006) both from the analytical standpoint (by accounting for regimes

of varying background mortality and colony size and by considering a broader class of examples in which aposematic traits are related implicitly) and from the practical standpoint (by assessing its efficacy and limitations in predicting the evolution of prey traits in finite simulated populations). Both developments constitute new contributions to the theory of aposematic signalling that are worth noting and the reader is encouraged to consult Scaramangas and Broom (2022) and Scaramangas et al. (2022) alongside this thesis.

Finally, we have provided two approaches that extend the original game to account for populations of prey population that consist of two types. While in the penultimate chapter the approach is general enough to account for a potentially large breadth of co-existence regimes the focus of that and the last chapter is on mimicry complexes. In these an undefended type known as the mimic resembles a defended type, known as the model so as to gain protection from predators. While there is ample empirical evidence to suggest that individuals from one species may gain selective advantage by resembling individuals from another, the mathematical modelling of Batesian mimicry is rather limited. We predict that models and mimics can co-exist along a continuum of solutions (representing the conspicuousness, noxiousness, and average mimic-to-model proportion) that are both ecologically and (locally) evolutionarily stable. While the approach of the penultimate chapter may be better suited to the modelling of automimicry the latter is more general and could better describe strict forms of (Batesian) mimicry. In the penultimate chapter we observe that that higher levels of conspicuousness are typically associated with higher levels of the defence and a lower mimetic load. Owing to the complexity of the presentation it was not possible to explore the impact of varying model defence on the continuum in the last chapter. Nonetheless, it was observed that the success of the mimic depends critically on the response of the predator and that this is likelier when the model is sufficiently aversive or that the relationship between the "mimetic load" and the equilibrium conspicuousness is an increasing one. The sensitivity of stability on the noxiousness of the model is perhaps a new an exciting development in the theory of Batesian mimicry.

If resources were ample we would seek to explore mimicry on three levels. The first would be to explore different sets of parameter values for the given example of chapter 6; the second would be to consider a larger breadth of examples through which to more clearly establish patterns of stability within the extended strategy space that include model defence. Third, would be to test our hypotheses against a more generalised genetic algorithm model that can account for evolution of traits in prey populations that consist of a model and a mimic (more general co-existence regimes may also be of interest to examine). It has been argued that controlling all parameters of the model (even in the single-species description) is a perhaps an unrealistic task and therefore the validation of stability predictions through genetic algorithm simulations are important especially since these rely on the assumption that prey populations are infinite. Much like for intermediate-sized populations of a single species we examined the roles of marginal mutant fitness, absolute resident fitness and stochasticity in the evolution of prey traits it would likewise be of interest to better understand the evolution prey traits of an intermediate-sized mimetic complex. On a separate note, it would also be of interest to incorporate aposematic colouration back into the single-type description and to examine the sensitivity of our more recent findings to the effects of kin grouping through non-point solutions.

Bibliography

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (Vol. 55). US Government printing office.

Abrams, P. A., Matsuda, H., & Harada, Y. (1993). Evolutionarily unstable fitness maxima and stable fitness minima of continuous traits. *Evolutionary Ecology*, *7*, 465–487.

Alcock, J. (1970). Punishment levels and the response of white-throated sparrows (zonotrichia albicollis) to three kinds of artificial models and mimics. *Animal Behaviour*, *18*, 733–739.

Allen, J., & Cooper, J. (1985). Crypsis and masquerade. *Journal of Biological Education*, *19*(4), 268–270.

Apaloo, J., Brown, J. S., & Vincent, T. L. (2009). Evolutionary game theory: Ess, convergence stability, and nis. *Evolutionary Ecology Research*, *11*(4), 489–515.

Arenas, L. M., Walter, D., & Stevens, M. (2015). Signal honesty and predation risk among a closely related group of aposematic species. *Scientific reports*, *5*(1), 1–12.

Argasinski, K. (2006). Dynamic multipopulation and density dependent evolutionary games related to replicator dynamics. a metasimplex concept. *Mathematical biosciences*, *202*(1), 88–114.

Argasinski, K., & Broom, M. (2018). Interaction rates, vital rates, background fitness and replicator dynamics: How to embed evolutionary game structure into realistic population dynamics. *Theory in Biosciences*, *137*(1), 33–50.

Balogh, A. C., & Leimar, O. (2005). Müllerian mimicry: An examination of fisher's theory of gradual evolutionary change. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1578), 2269–2275.

Barbaro, L., & Battisti, A. (2011). Birds as predators of the pine processionary moth (lepidoptera: Notodontidae). *Biological control*, *56*(2), 107–114.

Barnett, J. B., Michalis, C., Scott-Samuel, N. E., & Cuthill, I. C. (2018). Distance-dependent defensive coloration in the poison frog dendrobates tinctorius, dendrobatidae. *Proceedings of the National Academy of Sciences*, *115*(25), 6416–6421.

Bates, H. W. (1862). Contributions to an insect fauna of the amazon valley. lepidoptera: Heliconidae. *Transactions of the Linnean Society of London*, *23*, 495–566.

Baxandall, P. R., & Liebeck, H. (1986). *Vector calculus*. Oxford University Press, USA.

Bezzerides, A. L., McGraw, K. J., Parker, R. S., & Husseini, J. (2007). Elytra color as a signal of chemical defense in the asian ladybird beetle harmonia axyridis. *Behavioral Ecology and Sociobiology*, *61*(9), 1401–1408.

Bishop, D., & Cannings, C. (1976). Models of animal conflict. *Adv. Appl. Probl.*, *8*, 616–621.

Blount, J. D., Rowland, H. M., Drijfhout, F. P., Endler, J. A., Inger, R., Sloggett, J. J., Hurst, G. D., Hodgson, D. J., & Speed, M. P. (2012). How the ladybird got its spots: Effects of resource limitation on the honesty of aposematic signals. *Functional Ecology*, *26*(2), 334–342.

Blount, J. D., Speed, M. P., Ruxton, G. D., & Stephens, P. A. (2009). Warning displays may function as honest signals of toxicity. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1658), 871–877.

Bomze, I., & Pötscher, B. (1989). *Game Theoretical Foundations of Evolutionary Stability*. Springer-Verlag.

Boots, M., & Best, A. (2018). The evolution of constitutive and induced defences to infectious disease. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1883), 20180658.

Boyd, R., & Richerson, P. (2002). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, *215*(3), 287–296.

Brodie Jr, E. D. (1981). Phenological relationships of model and mimic salamanders. *Evolution*, 988–994.

Broom, M., Speed, M., & Ruxton, G. (2006). Evolutionarily stable defence and signalling of that defence. *Journal of theoretical biology*, *242*(1), 32–43.

Broom, M., Higginson, A. D., & Ruxton, G. D. (2010). Optimal investment across different aspects of anti-predator defences. *Journal of Theoretical Biology*, *263*(4), 579–586. https://doi.org/10.1016/j.jtbi.2010.01.002

Broom, M., Ruxton, G. D., & Schaefer, H. M. (2013). Signal verification can promote reliable signalling. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1771), 20131560.

Broom, M., Ruxton, G. D., & Speed, M. P. (2008). Evolutionarily stable investment in anti-predatory defences and aposematic signalling. *Mathematical modeling of biological systems*, *2*, 37–48.

Broom, M., & Rychtár, J. (2013). *Game-theoretical models in biology*. CRC Press.

Broom, M., Speed, M. P., & Ruxton, G. D. (2005). Evolutionarily stable investment in secondary defences. *Functional Ecology*, *19*(5), 836–843.

Brower, J. V. Z. (1958). Experimental studies of mimicry in some north american butterflies: Part ii. battus philenor and papilio troilus, p. polyxenes and p. glaucus. *Evolution*, 123–136.

Brower, J. v. Z. (1958a). Experimental studies of mimicry in some north american butterflies: Part i. the monarch, danaus plexippus, and viceroy, limenitis archippus archippus. *Evolution*, 32–47.

Brower, J. v. Z. (1958b). Experimental studies of mimicry in some north american butterflies. part iii. danaus gilippus berenice and limenitis archippus floridensis. *Evolution*, 273–285.

Brower, J. v. Z. (1960). Experimental studies of mimicry. iv. the reactions of starlings to different proportions of models and mimics. *The American Naturalist*, *94*(877), 271–282.

Brower, L. P., Brower, V. J., & Corvino, J. M. (1967). Plant poisons in a terrestrial food chain. *Proceedings of the National Academy of Sciences of the United States of America*, *57*(4), 893.

Brower, L. P. (1969). Ecological chemistry. *Scientific American*, *220*(2), 22–29.

Brower, L. P., Brower, J. V. Z., Stiles, F., Croze, H., & Hower, A. (1964). Mimicry: Differential advantage of color patterns in the natural environment. *Science*, *144*(3615), 183–185.

Brower, L. P., Cook, L. M., & Croze, H. J. (1967). Predator responses to artificial batesian mimics released in a neotropical environment. *Evolution*, 11–23.

Cannings, C., & Vickers, G. (1988). Patterns of ESS's 2. *Journal of Theoretical Biology*, *132*, 409–420.

Charlesworth, D., & Charlesworth, B. (1975a). Theoretical genetics of batesian mimicry i. single-locus models. *Journal of Theoretical Biology*, *55*(2), 283–303.

Charlesworth, D., & Charlesworth, B. (1975b). Theoretical genetics of batesian mimicry ii. evolution of supergenes. *Journal of Theoretical Biology*, *55*(2), 305–324.

Charlesworth, D., & Charlesworth, B. (1975c). Theoretical genetics of batesian mimicry iii. evolution of dominance. *Journal of Theoretical Biology*, *55*(2), 325–337.

Clarke, C., & Sheppard, P. M. (1975). The genetics of the mimetic butterfly hypolimnas bolina (l.) *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *272*(917), 229–265.

Cole, L. C. (1946). A theory of analyzing contagiously distributed populations. *Ecology*, *27*(4), 329–341.

Cook, L. M., Brower, L. P., & Alcock, J. (1969). An attempt to verify mimetic advantage in a neotropical environment. *Evolution*, 339–345.

Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., & Knuth, D. E. (1996). On the lambertw function. *Advances in Computational mathematics*, *5*(1), 329–359.

Cortesi, F., & Cheney, K. (2010). Conspicuousness is correlated with toxicity in marine opisthobranchs. *Journal of Evolutionary Biology*, *23*(7), 1509–1518.

Dahl, J., & Peckarsky, B. L. (2003). Does living in streams with fish involve a cost of induced morphological defences? *Canadian journal of zoology*, *81*(11), 1825–1828.

Daly, J. W. (2003). Ernest guenther award in chemistry of natural products. amphibian skin: A remarkable source of biologically active arthropod alkaloids. *Journal of medicinal chemistry*, *46*(4), 445–452.

Daly, J. W., & Myers, C. W. (1967). Toxicity of panamanian poison frogs (dendrobates): Some biological and chemical aspects. *Science*, *156*(3777), 970–973.

Darst, C. R., & Cummings, M. E. (2006). Predator learning favours mimicry of a less-toxic model in poison frogs. *Nature*, *440*(7081), 208–211.

Darst, C. R., Cummings, M. E., & Cannatella, D. C. (2006). A mechanism for diversity in warning signals: Conspicuousness versus toxicity in poison frogs. *Proceedings of the National Academy of Sciences*, *103*(15), 5852–5857.

Darst, C. R., Menéndez-Guerrero, P. A., Coloma, L. A., & Cannatella, D. C. (2005). Evolution of dietary specialization and chemical defense in poison frogs (dendrobatidae): A comparative analysis. *The American Naturalist*, *165*(1), 56–69.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Dawkins, R. (1999). *The extended phenotype: The long reach of the gene*. Oxford University Press, USA.

Deshmukh, R., Baral, S., Gandhimathi, A., Kuwalekar, M., & Kunte, K. (2018). Mimicry in butterflies: Co-option and a bag of magnificent developmental genetic tricks. *Wiley Interdisciplinary Reviews: Developmental Biology*, *7*(1), e291.

Dickinson, A., Boakes, R. A., & Sjödén, P.-O. (1981). Mechanisms of learning and motivation: A memorial volume to jerzy konorski.

Edmunds, J., & Edmunds, M. (1974). Polymorphic mimicry and natural selection: A reappraisal. *Evolution*, 402–407.

Edmunds, M., & Reader, T. (2014). Evidence for batesian mimicry in a polymorphic hoverfly. *Evolution*, *68*(3), 827–839.

Edwards, A. (2000). *Foundations of mathematical genetics*. Cambridge University Press.

Eshel, I. (1983). Evolutionary and continuous stability. *Journal of theoretical Biology*, *103*(1), 99–111.

Eshel, I. (1996). On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *Journal of mathematical biology*, *34*, 485–510.

Eshel, I., & Motro, U. (1981). Kin selection and strong evolutionary stability of mutual help. *Theoretical population biology*, *19*(3), 420–433.

Fisher, R. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.

Franks, D. W., Ruxton, G. D., & Sherratt, T. N. (2009). Warning signals evolve to disengage batesian mimics. *Evolution: International Journal of Organic Evolution*, *63*(1), 256–267.

Friedman, D. (1998). On economic applications of evolutionary game theory. *Journal of evolutionary economics*, *8*(1), 15–43.

Fryer, J. C. F. (1914). Vii. an investigation by pedigree breeding into the polymorphism of papilio polytes, linn. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *204*(303-312), 227–254.

Gabritschevsky, E. (1924). Farbenpolymorphismus und vererbung mimetischer varietäten der fliege volucella bombylans und anderer „hummelähnlicher "zweiflügler.

Gamberale-Stille, G., & Guilford, T. (2004). Automimicry destabilizes aposematism: Predator sample-and-reject behaviour may provide a solution. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1557), 2621–2625.

Gardner, M. (1970). Mathematical games: The fantastic combinations of John Conway's new solitaire game "life". *Scientific American*, *223*(4), 120–123.

Gols, R. (2014). Direct and indirect chemical defences against insects in a multitrophic framework. *Plant, Cell & Environment*, *37*(8), 1741–1752.

Haigh, J. (1975). Game theory and evolution. *Adv. Appl. Prob.*, *7*, 8–11.

Halpin, C. G., Skelhorn, J., & Rowe, C. (2008). Being conspicuous and defended: Selective benefits for the individual. *Behavioral Ecology*, *19*(5), 1012–1017.

Hämäläinen, L., Mappes, J., Thorogood, R., Valkonen, J. K., Karttunen, K., Salmi, T., & Rowland, H. M. (2020). Predators' consumption of unpalatable prey does not vary as a function of bitter taste perception. *Behavioral Ecology*, *31*(2), 383–392.

Hamilton, W. (1964). The genetical evolution of social behaviour. I and II. *Journal of Theoretical Biology*, *7*(1), 17–52.

Hamilton, W. (1967). Extraordinary sex ratios. *Science*, *156*, 477–488.

Hammill, E., Rogers, A., & Beckerman, A. P. (2008). Costs, benefits and the evolution of inducible defences: A case study with daphnia pulex. *Journal of evolutionary biology*, *21*(3), 705–715.

Hardy, G. (1908). Mendelian proportions in a mixed population. *Science*, *28*, 49–50.

Harsanyi, J. (1966). A general theory of rational behavior in game situations. *Econometrica: Journal of the Econometric Society*, 613–634.

Hastad, J., Just, B., Lagarias, J. C., & Schnorr, C.-P. (1989). Polynomial time algorithms for finding integer relations among real numbers. *SIAM Journal on Computing*, *18*(5), 859–881.

Hecht, M. K., & Marien, D. (1956). The coral snake mimic problem: A reinterpretation. *Journal of Morphology*, *98*(2), 335–365.

Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press.

Holen, Ø. H. (2013). Disentangling taste and toxicity in aposematic prey. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1753), 20122588.

Holen, Ø. H., & Svennungsen, T. O. (2012). Aposematism and the handicap principle. *The American Naturalist*, *180*(5), 629–641.

Irie, T., & Iwasa, Y. (2005). Optimal growth pattern of defensive organs: The diversity of shell growth among mollusks. *The American Naturalist*, *165*(2), 238–249.

Jeffords, M., Sternburg, J., & Waldbauer, G. (1979). Batesian mimicry: Field demonstration of the survival value of pipevine swallowtail and monarch color patterns. *Evolution*, 275–286.

Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions, volume 2* (Vol. 289). John wiley & sons.

Jones, R., Davis, S., & Speed, M. (2013). Defence cheats can degrade protection of chemically defended prey. *Ethology*, *119*(1), 52–57.

Kalmár, L. (1928). Zur theorie der abstrakten spiele. *Acta Sci. Math. Univ. Szeged*, *4*, 65–85.

Kato, H., & Takada, T. (2019). Stability and bifurcation analysis of a ratio-dependent community dynamics model on batesian mimicry. *Journal of Mathematical Biology*, *79*(1), 329–368.

Kauppinen, J., & Mappes, J. (2003). Why are wasps so intimidating: Field experiments on hunting dragonflies (odonata: Aeshna grandis). *Animal Behaviour*, *66*(3), 505–511.

Kikuchi, D. W., Barfield, M., Herberstein, M. E., Mappes, J., & Holt, R. D. (2022). The effect of predator population dynamics on batesian mimicry complexes. *The American Naturalist*, *199*(3), 000–000.

Kikuchi, D. W., & Pfennig, D. W. (2010). Predator cognition permits imperfect coral snake mimicry. *The American Naturalist*, *176*(6), 830–834.

Kikuchi, D. W., & Pfennig, D. W. (2012). A batesian mimic and its model share color production mechanisms. *Current Zoology*, *58*(4), 658–667.

Kirby, W., & Spence, W. (1817). An introduction to entomology. *An Introduction to Entomology, Volume 2*, *2*.

Kristiansen, E. B., Finkbeiner, S. D., Hill, R. I., Prusa, L., & Mullen, S. P. (2018). Testing the adaptive hypothesis of batesian mimicry among hybridizing north american admiral butterflies. *Evolution*, *72*(7), 1436–1448.

Kuchta, S. R., Krakauer, A. H., & Sinervo, B. (2008). Why does the yellow-eyed ensatina have yellow eyes? batesian mimicry of pacific newts (genus taricha) by the salamander ensatina eschscholtzii xanthoptica. *Evolution: International Journal of Organic Evolution*, *62*(4), 984–990.

Lee, T. J., Speed, M. P., & Stephens, P. A. (2011). Honest signaling and the uses of prey coloration. *The American Naturalist*, *178*(1), E1–E9.

Leimar, O., Enquist, M., & Sillen-Tullberg, B. (1986). Evolutionary stability of aposematic coloration and prey unprofitability: A theoretical analysis. *The American Naturalist*, *128*(4), 469–490.

Lewontin, R. (1961). Evolution and the theory of games. *Journal of theoretical biology*, *1*(3), 382–403.

Lieberman, E., Hauert, C., & Nowak, M. (2005). Evolutionary dynamics on graphs. *Nature*, *433*(7023), 312–316.

Lindstedt, C., Huttunen, H., Kakko, M., & Mappes, J. (2011). Disengtangling the evolution of weak warning signals: High detection risk and low production costs of chemical defences in gregarious pine sawfly larvae. *Evolutionary Ecology*, *25*(5), 1029–1046.

Lindström, L., Alatalo, R. V., & Mappes, J. (1997). Imperfect batesian mimicry—the effects of the frequency and the distastefulness of the model. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *264*(1379), 149–153.

Lotka, A. (1925). *Elements of physical biology*. Williams & Wilkins company.

Maan, M. E., & Cummings, M. E. (2012a). Poison frog colors are honest signals of toxicity, particularly for bird predators. *The American Naturalist*, *179*(1), E1–E14.

Maan, M. E., & Cummings, M. E. (2012b). Poison frog colors are honest signals of toxicity, particularly for bird predators. *The American Naturalist*, *179*(1), E1–E14.

Mallet, J. (2001). Mimicry: An interface between psychology and evolution. *Proceedings of the National Academy of Sciences*, *98*(16), 8928–8930.

Mappes, J., Marples, N., & Endler, J. A. (2005). The complex business of survival by aposematism. *Trends in ecology & evolution*, *20*(11), 598–603.

Martin, A. C., Zim, H. S., & Nelson, A. L. (1961). *American wildlife & plants: A guide to wildlife food habits: The use of trees, shrubs, weeds, and herbs by birds and mammals of the united states*. Courier Corporation.

Matessi, C., & Cori, R. (1972). Models of population genetics of batesian mimicry. *Theoretical population biology*, *3*(1), 41–68.

Matessi, C., & Gimelfarb, A. (2006). Discrete polymorphisms due to disruptive selection on a continuous trait—i: The one-locus case. *Theoretical Population Biology*, *69*(3), 283–295.

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.

Maynard Smith, J., & Parker, G. (1976). The logic of asymmetric contests. *Animal Behaviour*, *24*, 159–175.

Maynard Smith, J., & Price, G. (1973). The logic of animal conflict. *Nature*, *246*, 15–18.

McLean, I. (2011). *The relationship between chemical defence and death feigning in the red flour beetle (tribolium castaneum)* (Doctoral dissertation). Carleton University.

Merilaita, S., & Ruxton, G. D. (2007). Aposematic signals and the relationship between conspicuousness and distinctiveness. *Journal of Theoretical Biology*, *245*(2), 268–277.

Metz, J. A., Geritz, S. A., Meszéna, G., Jacobs, F. J., & Van Heerwaarden, J. S. (1995). Adaptive dynamics: A geometrical study of the consequences of nearly faithful reproduction.

Moran, P. (1958). Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, *54*(01), 60–71.

Mostler, G. (1935). Beobachtungen zur frage der wespenmimikry. *Zeitschrift für Morphologie und Ökologie der Tiere*, *29*(3), 381–454.

Müller, J. F. (1879). Ituna and thyridia; a remarkable case of mimicry in butterflies. *Transactions of the Entomological Society of London*, xx–xxix.

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, *36*(1), 48–49.

Nash, J. (1951). Non-cooperative games. *The Annals of Mathematics*, *54*(2), 286–295.

Nowak, M. (2006). *Evolutionary dynamics, exploring the equations of life*. Harward University Press, Cambridge, Mass.

Oaten, A., Pearce, C., & Smyth, M. (1975). Batesian mimicry and signal detection theory. *Bulletin of Mathematical Biology*, *37*(4), 367–387.

Palm, G. (1984). Evolutionary stable strategies and game dynamics for n-person games. *Journal of Mathematical Biology*, *19*(3), 329–334.

Pekár, S., Petráková, L., Bulbert, M. W., Whiting, M. J., & Herberstein, M. E. (2017). The golden mimicry complex uses a wide spectrum of defence to deter a community of predators. *Elife*, *6*, e22089.

Pfennig, D. W., Harcombe, W. R., & Pfennig, K. S. (2001). Frequency-dependent batesian mimicry. *Nature*, *410*(6826), 323–323.

Pfennig, D. W., Harper, G. R., Brumo, A. F., Harcombe, W. R., & Pfennig, K. S. (2007). Population differences in predation on batesian mimics in allopatry with their model: Selection against mimics is strongest when they are common. *Behavioral Ecology and Sociobiology*, *61*(4), 505–511.

Platt, A. P., Coppinger, R. P., & Brower, L. P. (1971). Demonstration of the selective advantage of mimetic limenitis butterflies presented to caged avian predators. *Evolution*, *25*(4), 692–701.

Pointer, M. D., Gage, M. J., & Spurgin, L. G. (2021). Tribolium beetles as a model system in evolution and ecology. *Heredity*, *126*(6), 869–883.

Poulton, E. B. (1890). *The colours of animals: Their meaning and use, especially considered in the case of insects*. D. Appleton.

Punnett, R. C. (1915). *Mimicry in butterflies*. University of Cambridge.

Rojas, B., Nokelainen, O., & Valkonen, J. K. (2021). Aposematism. In *Encyclopedia of evolutionary psychological science* (pp. 345–349). Springer.

Ross, S. M. (2019). *A first course in probability*. Pearson Boston.

Rothschild, M. (1963). Is the buff ermine (spilosoma lutea (huf.)) a mimic of the white ermine (spilosoma lubricipeda (l.))? *Proceedings of the Royal Entomological Society of London. Series A, General Entomology*, *38*(7-9), 159–164.

Rowe, C., & Guilford, T. (1999). The evolution of multimodal warning displays. *Evolutionary Ecology*, *13*(7), 655–671.

Rowland, H. M., Ruxton, G. D., & Skelhorn, J. (2013). Bitter taste enhances predatory biases against aggregations of prey with warning coloration. *Behavioral Ecology*, *24*(4), 942–948.

Roy, D. (1997). Communication signals and sexual selection in amphibians. *Current science*, 923–927.

Ruxton, G., Allen, T., W.L.and Sherratt, & Speed, M. (2019). *Avoiding attack: The evolutionary ecology of crypsis, aposematism, and mimicry*. Oxford University Press, USA.

Ruxton, G. D., & Beauchamp, G. (2008). The application of genetic algorithms in behavioural ecology, illustrated with a model of anti-predator vigilance. *Journal of theoretical biology*, *250*(3), 435–448.

Ruxton, G. D., Speed, M. P., & Broom, M. (2009). Identifying the ecological conditions that select for intermediate levels of aposematic signalling. *Evolutionary ecology*, *23*(4), 491–501.

Salas, L. S., Etgen, G. J., & Hille, E. (2007). *Calculus: One and several varialbles*. Wiley Sons.

Santos, J. C., & Cannatella, D. C. (2011). Phenotypic integration emerges from aposematism and scale in poison frogs. *Proceedings of the national academy of sciences*, *108*(15), 6175–6180.

Santos, J. C., Coloma, L. A., & Cannatella, D. C. (2003). Multiple, recurring origins of aposematism and diet specialization in poison frogs. *Proceedings of the National Academy of Sciences*, *100*(22), 12792–12797.

Scaramangas, A., & Broom, M. (2022). Aposematic signalling in prey-predator systems: Determining evolutionary stability when prey populations consist of a single species. *submitted*.

Scaramangas, A., Broom, M., Ruxton, G. D., & Roùviere, A. (2022). Evolutionarily stable levels of aposematic defence in prey populations. *submitted*.

Selten, R. (1965). Spieltheoretische behandlung eines oligopolmodells mit nachfrageträgheit: Teil i: Bestimmung des dynamischen preisgleichgewichts. *Zeitschrift für die gesamte Staatswissenschaft/Journal of Institutional and Theoretical Economics*, *121*(2), 301–324.

Selten, R. (1975). A reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, *4*, 25–55.

Shudo, E., & Iwasa, Y. (2001). Inducible defense against pathogens and parasites: Optimal choice among multiple options. *Journal of Theoretical Biology*, *209*(2), 233–247.

Skelhorn, J. (2015). Masquerade. *Current Biology*, *25*(15), R643–R644.

Skelhorn, J., & Rowe, C. (2006). Prey palatability influences predator learning and memory. *Animal Behaviour*, *71*(5), 1111–1118.

Skelhorn, J., & Rowe, C. (2007). Automimic frequency influences the foraging decisions of avian predators on aposematic prey. *Animal Behaviour*, *74*(5), 1563–1572.

Skelhorn, J., Rowland, H. M., Speed, M. P., & Ruxton, G. D. (2010). Masquerade: Camouflage without crypsis. *Science*, *327*(5961), 51–51.

Speed, M., Ruxton, G. D., Mappes, J., & Sherratt, T. N. (2012). Why are defensive toxins so variable? an evolutionary perspective. *Biological Reviews*, *87*(4), 874–884.

Speed, M. P., & Franks, D. W. (2014). Antagonistic evolution in an aposematic predator–prey signaling system. *Evolution*, *68*(10), 2996–3007.

Speed, M. P., & Ruxton, G. D. (2005). Warning displays in spiny animals: One (more) evolutionary route to aposematism. *Evolution*, *59*(12), 2499–2508.

Speed, M. P., & Ruxton, G. D. (2007). How bright and how nasty: Explaining diversity in warning signal strength. *Evolution*, *61*(3), 623–635.

Speed, M. P., Ruxton, G. D., Blount, J. D., & Stephens, P. A. (2010). Diversification of honest signals in a predator–prey system. *Ecology Letters*, *13*(6), 744–753.

Summers, K., Speed, M., Blount, J., & Stuckert, A. (2015). Are aposematic signals honest? a review. *Journal of evolutionary biology*, *28*(9), 1583–1599.

Summers, K., & Clough, M. E. (2001a). The evolution of coloration and toxicity in the poison frog family (dendrobatidae). *Proceedings of the National Academy of Sciences*, *98*(11), 6227–6232.

Summers, K., & Clough, M. E. (2001b). The evolution of coloration and toxicity in the poison frog family (dendrobatidae). *Proceedings of the National Academy of Sciences*, *98*(11), 6227–6232.

Svennungsen, T. O., & Holen, Ø. H. (2007). The evolutionary stability of automimicry. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1621), 2055–2063.

Tarvin, R. D., Borghese, C. M., Sachs, W., Santos, J. C., Lu, Y., O'connell, L. A., Cannatella, D. C., Harris, R. A., & Zakon, H. H. (2017). Interacting amino acid replacements allow poison frogs to evolve epibatidine resistance. *Science*, *357*(6357), 1261–1266.

Taylor, L. (1984). Assessing and interpreting the spatial distributions of insect populations. *Annual review of entomology*, *29*(1), 321–357.

Teichmann, J., Broom, M., & Alonso, E. (2014a). The evolutionary dynamics of aposematism: A numerical analysis of co-evolution in finite populations. *Mathematical Modelling of Natural Phenomena*, *9*(3), 148–164. https://doi.org/10.1051/mmnp/20149310

Teichmann, J., Broom, M., & Alonso, E. (2014b). The evolutionary dynamics of aposematism: A numerical analysis of co-evolution in finite populations. *Mathematical Modelling of Natural Phenomena*, *9*(3), 148–164. https://doi.org/10.1051/mmnp/20149310

Teichmann, J., Alonso, E., & Broom, M. (2015). A reward-driven model of darwinian fitness. *2015 7th International Joint Conference on Computational Intelligence (IJCCI)*, *1*, 174–179.

Trivers, R. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.

Tucker, A. (1980). On Jargon: The Prisoner's Dilemma. *UMAP Journal*, *1*(101).

Turner, J. R., KEARNEY, E. P., & EXTON, L. S. (1984). Mimicry and the monte carlo predator: The palatability spectrum, and the origins of mimicry. *Biological journal of the Linnean Society*, *23*(2-3), 247–268.

Twomey, E., Yeager, J., Brown, J. L., Morales, V., Cummings, M., & Summers, K. (2013). Phenotypic and genetic divergence among poison frog populations in a mimetic radiation. *PloS one*, *8*(2), e55443.

Vences, M., Kosuch, J., Boistel, R., Haddad, C. F., La Marca, E., Lötters, S., & Veith, M. (2003). Convergent evolution of aposematic coloration in neotropical poison frogs: A molecular phylogenetic perspective. *Organisms Diversity & Evolution*, *3*(3), 215–226.

Vickers, G., & Cannings, C. (1988). Patterns of ESS's 1. *Journal of Theoretical Biology*, *132*, 387–408.

Vidal-Cordero, J., Moreno-Rueda, G., López-Orta, A., Marfil-Daza, C., Ros-Santaella, J., & Ortiz-Sanchez, F. (2012). Brighter-colored paper wasps (polistes dominula) have larger poison glands. *Frontiers in zoology*, *9*, 20.

Vincent, T. L., & Brown, J. S. (2005). *Evolutionary game theory, natural selection, and darwinian dynamics*. Cambridge University Press.

Vincent, T. L., Cohen, Y., & Brown, J. S. (1993). Evolution via strategy dynamics. *Theoretical Population Biology*, *44*(2), 149–176.

Volterra, V. (1926). Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Mem. Accad. Sci. Lincei*, *2*, 31–113.

von Neumann, J. (1928). Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, *100*(1), 295–320.

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press, Princeton, NJ.

Waldbauer, G., & LaBerge, W. (1985). Phenological relationships of wasps, bumblebees, their mimics and insectivorous birs in northern michigan. *Ecological Entomology*, *10*(1), 99–110.

Wallace, A. R. (1877). The colors of animals and plants. *The American Naturalist*, *11*(11), 641–662.

Wallace, A. R. (1889). *Darwinism-an exposition of the theory of natural selection with some of its applications*. MacMillan & Co, London.

Wang, I. J. (2011). Inversely related aposematic traits: Reduced conspicuousness evolves with increased toxicity in a polymorphic poison-dart frog. *Evolution: International Journal of Organic Evolution*, *65*(6), 1637–1649.

Wang, L., & Broom, M. (2019). A theory for investment across defences triggered at different stages of a predator-prey encounter. *Journal of Theoretical Biology*, *473*, 9–19.

Waxman, D., & Gavrilets, S. (2005). 20 questions on adaptive dynamics. *Journal of evolutionary biology*, *18*(5), 1139–1154.

Weinberg, W. (1908). On the demonstration of heredity in man. *Papers on Human Genetics*, 4–15.

WolframAlpha. (2022). Nearest integer function. https://mathworld.wolfram.com/NearestIntegerFunction.html

Wright, S. (1930). Evolution in Mendelian Populations. *Genetics*, *16*, 97–159.

Yack, J. E. (2022). Acoustic defence strategies in caterpillars. In *Caterpillars in the middle* (pp. 195–223). Springer.

Zakharova, L., Meyer, K., & Seifan, M. (2019). Trait-based modelling in ecology: A review of two decades of research. *Ecological Modelling*, *407*, 108703.

Zermelo, E. (1913). Über eine anwendung der mengenlehre auf die theorie des schachspiels. *Proceedings of the Fifth International Congress of Mathematicians*, *2*, 501–504.

Zvereva, E. L., & Kozlov, M. V. (2016). The costs and effectiveness of chemical defenses in herbivorous insects: A meta-analysis. *Ecological Monographs*, *86*(1), 107–124.