The good, the finite, and the infinite

The good, the finite, and the infinite

Chai Molina, B.A., M.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY (2016) McMaster University
(Mathematics & Statistics) Hamilton, Ontario

TITLE: The good, the finite, and the infinite

AUTHOR: Chai Molina, B.A. (The Technion), M.Sc. (Tel Aviv University)

SUPERVISOR: Prof. David J. D. Earn

NUMBER OF PAGES: [xi], 193

ABSTRACT

Many interesting behaviours in the animal and human world involve cooperation among individuals. Yet, cooperating individuals are often susceptible to exploitation by cheaters. Because cheaters do better than the cooperators they exploit, the evolution and persistence of cooperation has been a challenging topic of study in biology, sociology and economics.

Studies often abstract from real cooperative interactions, and construct simple games in which players can choose either cooperation with other players, or defection, *e.g.,* the well known prisoner's dilemma and the snowdrift game. In these games and other social dilemmas, mutual cooperation yields greater payoffs than mutual defection, but individuals are still tempted to defect (because of the possibility that if they cooperate, the other player will defect).

Similar dilemmas also arise in situations where multiple individuals may be affected by the actions of one (such as volunteering for community service or evading taxes), and the main theme of this thesis is cooperation in groups. In chapter 2, we analyze pre-emptive vaccination for an outbreak of smallpox (following a bioterrorist attack or accidental release), from the public health (*i.e.,* group) and individual perspectives. Chapters 3 and 4 deal with an extension of the snowdrift game to n interacting players and continuous strategy sets (where individuals decide on their degree of cooperation): in chapter 3, we analyze global evolutionary stability of cooperative strategies in a large class of n-player snowdrift games in infinite populations; chapter 4 analyzes general continuous n-person snowdrift games in finite populations, and compares the evolutionary dynamics with their infinite population analogues. In chapter 5, we present a general framework to model selection processes in finite populations, necessary for the analysis in chapter 4.

ACKNOWLEDGEMENTS

Countless times in the past fifty four months, I was reminded that choosing to pursue my PhD at McMaster University was the right decision. While this decision could be justified by the many wonderful people I've had the privilege of meeting, or experiences I've had as a result of this choice, the primary justification is Prof. David Earn. I simply cannot imagine a better mentor. Within academia or elsewhere, I have never seen anyone as invested as David is in their mentees' intellectual, professional and personal development and well being, and that makes all the difference. His curiosity, attention to detail, breadth of knowledge and depth of understanding are phenomenal, and I would be lucky to have absorbed even a fraction of these qualities. Beyond mentoring, teaching and guiding, he was a friend, and even a surrogate family at a time when mine were across the pond. Thank you.

The members of my supervisory committee, Profs. Sigal Balshine, Paul Higgs and Rufus Johnstone, were also instrumental in my academic development and the production of this thesis. In particular, I'd like to thank Sigal and Rufus for helping me connect the abstract equations to the real, living world, Rufus and Paul for many useful comments, and all three for insightful and enlightening discussions.

I have also benefited from interaction with, learning from and the contribution of other members of the McMaster mathematical biology community. I'd especially like to thank Profs. Ben Bolker and Jonathan Dushoff for their difficult questions, and the ensuing discussions. I thank also Profs. Eric Sawyer and Gail Wolkowicz for helping ease my way back into math after a long hiatus.

Many other members of the McMaster community have both directly and indirectly contributed to my PhD. I thank the EarnLab, David Champredon, Sarah Drohan, Karsten Hempel, Irena Papst and Dora Rosati, for everything from widening my perspective of mathematical biology to sitting through and providing helpful comments on practice talks, and many many laughs on meetings, trips to conferences and coffee breaks. A special thank you to my friend and colleague Lindsay Keegan, for helping convince me to come to McMaster, for her support and advice throughout these past years, as well as many fun adventures, parties, lunches, and likely a couple of weeks in coffee breaks. I am also indebted to my partner, Shannon Falconer, for her encouragement and help on the home stretch, as well as for putting things into perspective and saving me from sleep deprivation.

Last, but certainly not least, my gratitude goes also to my parents, Eva and Armand, and my sister Rotem. From encouraging my childhood self to pursue my curiosity and passion to learn, through advising me in my decision to study under Prof. Earn, supporting me when I was close to throwing in the towel, and much, much more, they were always there for me; many of the lessons they taught me earlier in life have served me well throughout this journey.

DECLARATION OF ACADEMIC ACHIEVEMENT

Each chapter of this thesis has been written as a separate manuscript. Chapter 2 has been published, chapter 3 has been revised and resubmitted for publication and chapters 4 and 5 are in the final stages of preparation for publication. Programming, analysis and manuscript preparation for each chapter was primarily an individual effort, with contributions in programming, analysis and editing from David Earn.

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

Cooperation is defined as "to associate with another or others for mutual benefit" [1], and encompasses some of the most interesting behaviours in the living world. For example, humans have comensal relationships with their skin and gut microbiomes [2, 3]; they cooperate with their spouses [4, 5], neighbours [6] and even their foes [7]; and nations cooperate in defense, economic and environmental agreements [8, 9, 10]. Thus, even restricting attention to interactions involving humans, one can still observe the vast range of scales throughout which examples of cooperation are found.

Yet, in many instances of cooperation, cooperating individuals may be exploited by cheaters who benefit from the cooperation of others [11, 12, 13, 14], without cooperating themselves. This leads to a fundamental problem: it has been said that "[n]othing in biology makes sense except in the light of evolution" [15]. Evolution by natural selection is one of the fundamental tenets of the modern evolutionary synthesis [16, 17], and is succinctly expressed as "[a]ll life evolves by the differential survival of replicating entities" [18]. Thus, because cheaters do better than the cooperators they exploit, Richard Dawkins wrote: "Be warned that if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature" [18]. Hence, biologists, sociologists, economists, political scientists and mathematicians have gone to great lengths to explain how cooperation can evolve and persist [19, 20, 21, 22, 23, 24, 25].

Theoretical and empirical studies often abstract the salient components of specific interactions among independent individuals in natural situations and construct idealized games in which players can choose whether or not to cooperate with other players, and sometimes also to what degree they wish to cooperate. Perhaps the most well-known of these games is the *"prisoner's dilemma"* [19, 26, 27], a symmetric two-player game with two strategies, cooperate or defect, represented by the payoff matrix in table 1.1.

It is assumed that $T > R > P > S$ where the variables T, R, P, and S denote the

strategy	Cooperate	Defect
Cooperate	R	S
Defect	T	P

Table 1.1: The payoff matrix for the Prisoner's Dilemma ($T > R > P > S$).

temptation to defect, the reward for cooperation, the punishment for mutual defection, and the "sucker's payoff" received for cooperating while the other player defects, respectively. The meaning of these conditions is as follows:

- $T > R$: If the opponent cooperates, the temptation to defect is larger than the reward for cooperation.
- $P > S$: If the opponent defects, the sucker's payoff is smaller than the punishment for mutual defection, so defecting is preferable to cooperation.
- $R > P$: The reward for mutual cooperation is larger than the punishment for mutual defection.

The first two conditions imply that regardless of its opponent's strategy, the focal player obtains a higher payoff by defecting than by cooperating (defection dominates cooperation) — and thus a rational player would always choose defection. The last condition ensures that if both players defect, they gain less than if they had both cooperated, which is the source of the "dilemma".

Note that in the iterated prisoner's dilemma (that is, when the game is played repeatedly between two opponents), the condition

$$2R > T + S\,, \tag{1.1}$$

is added in order to prevent two players who alternate between cooperation and defection asynchronously (so that at any iteration, one cooperates and the other defects) from obtaining a higher payoff than players who simply cooperate with one another.

Another common game that arises in studies of cooperation is the "snowdrift game" (also known as chicken, or the hawk-dove game)[28]. The snowdrift game is also described by a payoff matrix identical to table 1.1, except that the sucker's payoff is taken to be larger than the punishment for mutual defection ($S > P$), which facilitates the persistence of cooperation because defection no longer dominates cooperation [29, 30].

The condition that mutual cooperation yields greater payoffs than mutual defection is common to both the prisoner's dilemma and the snowdrift game, and is a fundamental aspect of social dilemmas [31]: if only everyone cooperated, they would be better off than when they all defect. In fact, from the perspective of the population condition (1.1) also implies that the average payoff is largest when everyone cooperates.

In the models of cooperation mentioned thus far (the prisoner's dilemma and the snowdrift game), individuals interact in pairs. However, there are many examples where it is more appropriate to model interactions among groups of individuals or at the population level, such as volunteering for community service [6], group vigilance behaviour [32, 33], and microbes secreting beneficial compounds which are then accessible to others [34, 35].

The main theme of this thesis is cooperation in groups. Chapter 2 concerns a specific example of population-level cooperation in humans: the vaccination game. The next two chapters deal with an extension of the snowdrift game to n interacting players and continuous strategy sets: in chapter 3, we analyze a large class of biologically sensible n-player snowdrift games in *infinite* populations (a mathematically convenient assumption often made in the literature); in chapter 4, we then analyze general continuous n-person snowdrift games in *finite* populations, and compare the resulting dynamics to those obtained in infinite populations. Lastly, chapter 5 concerns general selection processes in finite populations. While chapter 5 does not directly concern cooperation, its results are used in chapter 4. The four chapters and their main results are introduced in more detail below.

1.1 The vaccination game

One important example of population-level cooperative interaction is the "vaccination game" [36, 37, 38]. Individuals choose whether or not to vaccinate based on real or perceived costs of vaccination. When the vaccine coverage is high enough, those individuals who are still susceptible benefit from a significantly lower probability of contracting the disease as a result of the vaccinators' immunity [39]. Society relies on this effect, called "herd immunity" [40], to protect individuals who cannot be vaccinated [40, 41] (*e.g.,* infants, the elderly and immuno-compromised individuals), as well as those who the vaccine has not rendered immune (vaccines are rarely 100% effective [42, 43]). However, herd immunity is also susceptible to abuse by "freeriders", who choose to forgo vaccination for selfish reasons[44, 45].

Since the proportion of people who choose to vaccinate is typically smaller than the vaccine coverage that is best from the group perspective [36, 37], this leads to a question quite distinct from those regarding the evolution and persistence of cooperation: what is the cost (*e.g.,* in terms of increased mortality) incurred by the group due to freeriders causing a deviation from the group-optimal vaccine coverage? This was the main question studied by Bauch *et al.* [37]. In the scenario they analyzed, there is a certain probability of an outbreak of smallpox (caused by either accidental release or bioterrorism). Individuals can choose to vaccinate pre-emptively ("vaccinators"), delay and attempt to vaccinate after an outbreak has begun ("delayers"), or vaccinate preemptively with a probability p that is not 0 or 1 ("mixed strategists"). In the event of an outbreak, vaccination of susceptible individuals

was assumed to begin after 14 days, and then proceed at a constant rate of 10% of the total population per day, until no susceptible individuals remain.

The standard smallpox vaccination procedure amounts to infection with *vaccinia*, a virus related to *variola* (which causes smallpox). Vaccination with *vaccinia* carries relatively high risks of morbidity and mortality compared with vaccines commonly in use today [46] (which are still far lower than those of a smallpox infection). Thus, vaccinators risk morbidity and mortality associated with the vaccine, while delayers expose themselves to these risks only if an outbreak occurs, at the cost of the additional risk of contracting smallpox.

Bauch *et al.* [37] showed that this vaccination game has a convergently stable Nash equilibrium p_{i}: if the entire population plays the strategy p_{i}, an individual cannot obtain a higher survival probability than the remainder of the population by (unilaterally) changing their strategy. This ***individual equilibrium*** results in a vaccination coverage that is suboptimal from the group perspective, in that overall mortality is not minimized.

Chapter 2 expands on the analysis of Bauch *et al.* [37] by considering, beyond the constant-rate vaccination originally studied, four additional post-outbreak scenarios. Three of these scenarios take into account changes in the delayers' willingness to vaccinate based on media reports regarding the progression of the epidemic. A fourth model of instantaneous vaccination of a proportion of the population is also considered, because experts have suggested that vaccination of the entire US population is achievable in 3 days [47], which is much shorter than the latent period of smallpox, estimated to be 15 days [48]. In contrast to the constant-rate vaccination scenario of Bauch *et al.* [37], we were able to obtain some analytical results—including the individual equilibrium and final sizes—for most of these models.

Similar to the vaccination rate in the original model of Bauch *et al.* [37], the post-outbreak vaccination models we consider contain a numerical parameter, which we call ***vaccination effort***, with which the vaccination rate increases (for any given proportions of remaining susceptible and infective individuals). For a large class of biologically sensible models of post-outbreak vaccination (including, but not limited to, the five described above), we identified a ***mortality plateau***, that is, conditions under which increasing the vaccination effort does not reduce mortality, and explained the underlying cause for the plateau's existence. Any lag between the beginning of an outbreak and the initiation of the post-outbreak vaccination response extends the mortality plateau to higher vaccination efforts, thus making it harder to decrease mortality by increasing the vaccination effort. Moreover, because in reality there will be a maximal feasible vaccination effort (dependent on the capabilities of the public health authorities), when the lag between the beginning of the outbreak and the initiation of the vaccination response is large enough, it will be practically impossible to reduce mortality by increasing the vaccination effort. The effective critical lag, beyond which mortality is constant for all feasible vaccination efforts, depends on the post-outbreak vaccination model used. These results highlight both the

need for a quick post-outbreak response, and the importance of efforts to determine likely post-outbreak vaccination scenarios.

1.2 Evolutionary stability in nonlinear public goods games

A public good is a resource from which it is impossible to exclude others, and such that consumption by one individual does not reduce the amount available for another's consumption (although the latter property is sometimes relaxed). The theory of public goods offers insights into many problems in biology, from the formation of bacterial biofilms [49] and cancer [50, 51] to major evolutionary transitions [52] and even the evolution of life on earth [53] (see [54] for a review).

One classic public goods game is the continuous n-player snowdrift game [55], a natural extension of the 2-player version described earlier: players individually choose their level of contribution to a public good. Each player pays a cost associated with its contribution, but its payoff is a function of the combined contributions of all n interacting players.

Typically, evolutionary stability in public goods games is analyzed using frameworks that assume an infinite population (*e.g.,* adaptive dynamics [56, 57, 58, 59]). In an infinite population, a single invading mutant does not affect the mean fitness of individuals playing the resident strategy. However, there are a number of biologically sensible scenarios under which invading mutants might affect the residents' fitness. In small finite populations, a single invader's effect on the resident population is not necessarily negligible. Even in large populations, genetic drift, migration and environmental variability can all allow invading mutants to become a large enough proportion of the entire population so as to significantly affect the resident strategy's payoffs.

In chapter 3, we analyze a large, biologically interesting sub-class of n-player snowdrift games in infinite populations and find necessary and sufficient conditions for the existence of a cooperative globally evolutionarily stable strategy (ESS). We also show that the cooperative global ESS that we identify is locally evolutionarily and convergently stable in a much more general sense: when invaded by mutants comprising any proportion of invaders $0 < \epsilon < 1$, if the mutant strategy is close enough to the cooperative ESS, mutants will be selected against. Moreover, if a population playing a strategy sufficiently close to the cooperative ESS is invaded by a proportion of mutants $0 < \epsilon < 1$ playing a strategy between the resident strategy and the ESS, the mutants will be selected for.

1.3 Evolutionarily stability in continuous public goods games in finite populations

In chapter 4, we analyze evolutionary stability in general n-player snowdrift games in finite, well-mixed populations. We show that populations playing strategies expected to be evolutionarily stable based on infinite-population analyses can always be invaded and replaced by sufficiently close, less cooperative strategies. We then find conditions for local evolutionary and convergent stability in finite populations, and identify the reason for the discrepancy between ESSs in infinite and finite populations: on average, in an infinite population, mutants interact with more mutants and residents interact with fewer mutants than they would in a finite population.

We compare the evolutionary outcomes expected in finite and infinite populations in the sub-class of n-player snowdrift games analyzed in chapter 3. We find a sub-set of this class of n-player snowdrift games that—despite having a cooperative ESS when played in an infinite population—have no cooperative ESS when played in some finite populations. Importantly, we identify conditions under which no cooperative ESS exists (i) only for sufficiently small population sizes, (ii) for any sufficiently large population size, or (iii) for any finite population size. Thus, the qualitative difference in evolutionary outcomes between finite and infinite populations does not necessarily disappear as population size is increased.

1.4 On general models of selection in finite populations

In population genetics, one is typically interested in the dynamics of gene frequencies in a population over time [60, 61, 62]. In finite, well-mixed populations, the canonical choices for describing these dynamics are the Moran [63] and Wright-Fisher [64, 65] (WF) processes. However, these models are not applicable to all biological situations [66, 67, 68], and alternative models have been shown to exhibit very different qualitative behaviours [67, 68, 69, 70, 71, 72, 73, 74, 75], which motivated research on a class of generalized Wright-Fisher (GWF) processes [68, 75].

When analyzing the evolution of strategies in a population of agents playing games, one must decide how to translate payoffs into differential reproductive success, which is closely tied with the population-genetic problem described above. Almost all theoretical results in evolutionary game theory relate to either the Moran or WF process, but evidence that the underlying selection process can drastically affect evolutionary dynamics [76] has recently motivated the study of more general processes [77, 78, 79].

In chapter 5, we use the theories of Markov chains and martingales [80] to define and analyze a general class of mutationless selection processes in finite populations. These are discrete-time stochastic processes describing changes in the frequencies of two traits in a

population over time, with homogeneous populations (in which all individuals have the same trait) being absorbing (*i.e.,* if one type becomes extinct, it can never re-invade). A neutral drift process is then defined as a selection process that is additionally a martingale (that is, the mean trait frequencies do not change from one time-step to the next).

Being concerned with applications to evolutionary game theory in finite populations (in which we typically only require bounds on fixation probabilities, rather than fixation times and continuum limits) allows us to analyze a larger class than the GWF models. Beyond presenting a general framework in which to think about selection and neutral drift, our main result, lemma 5.5.4, extends a known theorem concerning the WF process [81, Theorem 1] to general selection processes: if one type is always as fit as the other (regardless of the population composition) and is strictly fitter at some population composition that can be reached from the initial one, then its fixation probability is higher than that expected in a neutral drift process. While simple and intuitive, this result is crucial to the analysis conducted in chapter 4.

Chapter 2

Game theory of pre-emptive vaccination before bioterrorism or accidental release of smallpox

Molina, C. and Earn, D. J. D. (2015). *Journal of The Royal Society Interface*, 12(107):2041387.
DOI: 10.1098/rsif.2014.1387

2.1 Abstract

Smallpox was eradicated in the 1970s, but new outbreaks could be seeded by bioterrorism or accidental release. Substantial vaccine-induced morbidity and mortality make pre-emptive mass vaccination controversial, and if vaccination is voluntary then there is a conflict between self- and group-interests. This conflict can be framed as a tragedy of the commons, in which herd immunity plays the role of the commons, and free-riding (*i.e.,* not vaccinating pre-emptively) is analogous to exploiting the commons. This game has been analyzed previously for a particular post-outbreak vaccination scenario. We consider several post-outbreak vaccination scenarios and compare the expected increase in mortality that results from voluntary *vs.* imposed vaccination. Below a threshold level of post-outbreak vaccination effort, expected mortality is independent of the level of response effort. A lag between an outbreak starting and a response being initiated increases the post-outbreak vaccination effort necessary to reduce mortality. For some post-outbreak vaccination scenarios, even modest response lags make it impractical to reduce mortality by increasing post-outbreak vaccination effort. In such situations, if decreasing the response lag is impossible, the only practical way to reduce mortality is to make the vaccine safer (greater post-outbreak vaccination effort leads only to fewer people vaccinating pre-emptively).

2.2 Introduction

The number of annual cases of smallpox in the early 1950's, just prior to the WHO global eradication program, is estimated at 50 million [46]. The eradication campaign was successful [46], but samples of the variola virus are still kept in at least two known laboratories in Russia and the United states [82]. In a worrying incident in July 2014, previously forgotten vials containing samples of smallpox, some of which were viable, were found in a lab at the National Institute of Health campus in Bethesda, MD [83]. Thus, the threat of the reintroduction of smallpox, whether inadvertently or in a bioterrorist attack, is still present.

Consequently, some countries—notably the United States—are interested in measures to protect their populations from potential smallpox infection. Prophylactic vaccination for smallpox carries a high cost (relative to other vaccines in use today), as the probability of death following vaccination—or "risk from being vaccinated"—is $r_{\mathrm{v}} \simeq 10^{-6}$ and serious side-effects occur with probability $\sim 10^{-3}$ [46]. Of course, infection with smallpox carries a much greater risk, since the case fatality proportion—the "risk from infection"—is $r_{\mathrm{i}} \simeq 0.3$ [46]. (See table 2.1 for a summary of parameter estimates.)

The substantial vaccine-induced morbidity and mortality associated with smallpox vaccination make pre-emptive mass vaccination controversial. If vaccination is voluntary, there is a conflict between self- and group-interests. This conflict can be framed as a tragedy of the commons, in which herd immunity plays the role of the commons, and free-riding (i.e., not vaccinating pre-emptively) is analogous to exploiting the commons. A previous game-theoretical study by Bauch and co-workers [37] examined this conflict of interest, and focused on the trade-off between prophylactic vaccination and post-outbreak mass vaccination (which has been shown to outperform contact-traced vaccination in a bioterrorism setting [84]). In particular, they showed that if the decision regarding pre-emptive vaccination is left to the individual, the vaccine coverage achieved will be sub-optimal from the group perspective. Bauch *et al.* [37] assumed that once a post-outbreak vaccination campaign begins, individuals will be vaccinated at a constant rate determined by existing infrastructure.

Various mechanisms might drive the rate of vaccination. Vaccination at a constant rate might be achieved if vaccination centres are flooded by individuals seeking the vaccine, and are operating at peak capacity. However, public responsiveness to such a campaign is hard to predict. If demand for the vaccine does not exceed the maximal rate of distribution by public health services, the post-outbreak dynamics might play out differently, depending on the public's reaction patterns. For example, media reports on the number of new cases might influence individuals to obtain the vaccine, in which case it is reasonable to model the vaccination rate as proportional to smallpox incidence.

In this paper, we return to the problem posed by Bauch *et al.* [37], but compare a variety of possible post-outbreak vaccination scenarios (described intuitively in § 2.3 and in precise mathematical terms in § 2.6). Whereas the scenario considered in [37] could

only be analyzed numerically, several of the vaccination scenarios that we consider here can be addressed analytically to obtain exact results. To this end, in § 2.4 we make some adjustments to the game-theoretical framework of Bauch and Earn [36] so that it can be applied to the scenarios we investigate here.

Throughout this paper we use smallpox as an illustrative example. However, our analyses can be applied to any vaccine-preventable infectious disease that could be used for bioterrorism or released accidentally, and for which the Susceptible-Infectious-Removed (SIR) or Susceptible-Exposed-Infectious-Removed (SEIR) models are applicable (see §2.6). Our qualitative results appear to be robust to which post-outbreak vaccination scenario is considered and the specific parameter values (we prove this in some cases), but the precise numerical values will vary.

We calculate the vaccination coverage obtained by voluntary pre-emptive vaccination and assess the costs of this policy as compared to mandatory vaccination. The group-optimal pre-emptive vaccine coverage is discussed in §2.5. We discuss parameter estimates and the procedure used to compare the the various models fairly in § 2.7. We compare the predictions of the various models, and emphasize important considerations for public health in §4.4. Notation and definitions are summarized in Tables 2.1, 2.2 and 2.3.

2.3 Vaccination scenarios

In this section, we give a brief description of the various post-outbreak vaccination scenarios considered in this paper.

Media coverage of a smallpox outbreak is likely to influence individual decisions concerning vaccination. Measures of severity of the outbreak that are likely to appear in the media include:

- Death rates, as in "300 people died of smallpox today",
- Total number of people currently infected (prevalence), as in "There are now 30,000 people sick with smallpox",
- New cases (incidence), as in "200 new cases of smallpox were confirmed today".

We consider separately how each of these types of information could affect smallpox vaccine uptake; in each case, we assume that the vaccination rate is proportional to the relevant quantity (*e.g.*, prevalence). Note that in standard epidemiological models [85], death rate is proportional to prevalence, so the first and second cases above are mathematically identical.

As a type of "null model" for media-induced vaccination, we also consider the situation in which

- vaccination rate is simply proportional to the size of the remaining susceptible population; this corresponds to a constant *per capita* vaccination rate for susceptible individuals (see appendix 2.B.1). This can be regarded as a "null model" to compare with models for the scenarios above in the following sense: Individuals' proclivity to vaccinate is constant over time, and does not depend on the state of the epidemic (*i.e.,* on prevalence or incidence, which are likely to be reported by the media), while the vaccination rate falls as the number of susceptibles decreases over time, meaning that fewer individuals per unit time are inclined to vaccinate.

We also consider two scenarios in which vaccine uptake is not influenced by the media, but is constrained by the capabilities of public health authorities:

- If an outbreak occurs, immediately vaccinate a proportion of the susceptible population. The proportion might describe the efficacy of a post-outbreak campaign in convincing those who have thus-far avoided vaccination. Individuals who remain unvaccinated after this post-outbreak campaign would be persons holding particularly radical anti-vaccine opinions.
- Susceptible individuals are vaccinated at a constant rate until there are no more susceptibles remaining.

Finally, for each of the above scenarios, we investigate the effect of a lag between the start of an outbreak and the initiation of the post-outbreak vaccination response (allowing for public health authorities to organize a response to the outbreak). Bauch *et al.* [37] assumed such a response lag in their model, which is otherwise identical to the final scenario described above.

The epidemic models associated with each of the above five scenarios are described in detail in § 2.6.

2.4 Game-theoretical formulation

In this section, we adapt the game-theoretical framework of Bauch and Earn [36] to our current problem. We assume that all individuals have full knowledge and are rational (in the game-theoretical sense; see [86]).

We denote the proportion of the population vaccinated pre-emptively as p. Because a proportion r_{v} of those vaccinated will die, the pre-outbreak vaccine coverage (the proportion of the population that is immune prior to the outbreak) is $p_{\mathrm{eff}} = p\frac{1-r_{\mathrm{v}}}{1-pr_{\mathrm{v}}}$ [37], which is slightly smaller than p. But, since none of the mathematical analysis and conclusions which follow are affected by this, and because the difference between p and p_{eff} is negligible, we refer to p as the pre-outbreak vaccine ***coverage level*** for simplicity (as in [37]).

Let $a \in [0, 1]$ be the probability of an outbreak ('a' for 'bioterrorist *attack* probability'

or '*accidental* release probability') per lifetime (or whatever time period is under consideration). Consider two pure strategies: ***vaccinate*** and ***delay***. The former vaccinates pre-emptively, before the beginning of an outbreak, and so receives (expected) payoff $-r_{\mathrm{v}}$; the latter delays vaccination until after an outbreak (at which point s/he may still be vaccinated during the public health post-outbreak vaccination campaign), and receives payoff

$$-a[r_{\mathrm{i}}\pi_p + \psi_p r_{\mathrm{v}}]\,, \tag{2.1}$$

where π_p and ψ_p are the probabilities of a delayer being infected, or vaccinated, respectively, after an outbreak (the delayer infection and vaccination probabilities are discussed in more detail in § 2.6). A ***mixed strategy*** is specified by the probability P that an individual will choose to vaccinate pre-emptively. We also assume $r_{\mathrm{v}} < ar_{\mathrm{i}}$ because if it were not so, even if all delayers were infected in an outbreak, the risk of dying in an outbreak would be smaller than the risk of dying from the side-effects of the vaccine, hence there would be no reason to vaccinate.

The payoff to an individual playing a mixed strategy (vaccinating with probability P) in a population in which the coverage level is p, is given by

$$E(P,p) = -Pr_{\mathrm{v}} - (1-P)a(\pi_p r_{\mathrm{i}} + \psi_p r_{\mathrm{v}})\,. \tag{2.2}$$

Equivalently, defining the ***relative risk*** of vaccination compared with infection as

$$r = \frac{r_{\mathrm{v}}}{r_{\mathrm{i}}}\,, \tag{2.3}$$

we have $E(P,p) = -r_{\mathrm{i}}[rP + (1-P)a(\pi_p + \psi_p r)]$. Since the parameter r_{i} simply scales the game payoff by a constant, it does not change the dynamics. We therefore use the rescaled payoff function

$$E(P,p) = -[rP + (1-P)a(\pi_p + \psi_p r)]\,. \tag{2.4}$$

Suppose that a proportion ϵ of the population vaccinate with probability P and $1-\epsilon$ vaccinate with probability Q. Following [36], we assume 100% vaccine efficacy, which implies coverage level $p = \epsilon P + (1-\epsilon)Q$. (Note that in a homogeneous population where all individuals play the same strategy P, *i.e.,* $\epsilon = 1$, the coverage is $p = P$.) The payoffs to individuals playing P and Q in such a population are then

$$E_P(P,Q,\epsilon) = E(P, \epsilon P + (1-\epsilon)Q) \tag{2.5a}$$

$$E_Q(P,Q,\epsilon) = E(Q, \epsilon P + (1-\epsilon)Q)\,, \tag{2.5b}$$

respectively, and the ***payoff gain*** to an individual playing P rather than Q in this population

is

$$\begin{aligned}\Delta E &= E_P(P,Q,\epsilon) - E_Q(P,Q,\epsilon) \\ &= -[rP + (1-P)a(\pi_p + \psi_p r)] + [rQ + (1-Q)a(\pi_p + \psi_p r)] \\ &= \left(\pi_p + r\psi_p - \frac{r}{a}\right)a(P-Q)\,, \qquad \text{where } p = \epsilon P + (1-\epsilon)Q. \end{aligned} \tag{2.6}$$

A strategy P^* is a ***Nash equilibrium*** (NE) if and only if (*iff*) in a population in which all individuals are playing P^*, no player employing a different strategy can achieve a higher payoff. Mathematically, this means that for any other strategy $Q \in [0,1]$ if the proportion playing Q is small enough (*i.e.,* $1-\epsilon$ is sufficiently small), then the payoff gain to strategy P^* is non-negative, *i.e.,* $\Delta E(P^*, Q, \epsilon) \geq 0$. When such a NE exists, we refer to this strategy as the ***individual equilibrium*** and denote it by p_{i}. This equilibrium is "individual" in the sense that it is determined by individuals attempting to maximize their payoffs (unlike the group optimum discussed in § 2.5 below). Note, however, that this is a ***population game*** [28, 36] so the payoff to individuals depends on the frequencies of strategies in the entire population.

Additionally, consider a scenario whereby strategy P invades a population playing strategy Q. If in this scenario, strategies P that are closer to the NE P^* than the prevalent strategy Q, obtain a higher payoff than the prevalent strategy, then P^* is called a ***convergently stable Nash equilibrium*** (CSNE). Mathematically, this is equivalent to demanding that if $\epsilon \ll 1$ then

$$P^* < P < Q \leq 1 \implies \Delta E(P,Q,\epsilon) \geq 0$$

and

$$0 \leq Q < P < P^* \implies \Delta E(P,Q,\epsilon) \geq 0.$$

In order to proceed with the analysis, it is necessary to derive the probabilities π_p and ψ_p from an epidemiological model, either numerically or analytically (see appendix 2.E). Proofs of existence and uniqueness of a CSNE are given for several cases in appendix 2.G. These proofs depend on π_p being a decreasing function of p. We have shown this to be true when post-outbreak vaccination is instantaneous or proportional to incidence, and also when vaccination is proportional to prevalence and $\alpha\phi_{\text{prev}} > \gamma(1-\alpha)$. Based on biological intuition corroborated with simulations, we have assumed that π_p decreases with p for all the models considered here. This has also recently been proved for other post-outbreak vaccination models not considered here [F. Bai and F. Brauer, pers. comm.].

2.5 Group optimum

From the perspective of a public health official (*i.e.,* group interest), it is desirable to attain the vaccine coverage that minimizes mortality. From this group perspective, a strategy is specified by the proportion p of the population that is pre-emptively vaccinated. The currency with which we compare strategies is the ***mortality cost*** $C(p)$, *i.e.,* the proportion of the population that is expected to die (either from smallpox infection or vaccination),

$$C(p) = rp + (1-p)a(\pi_p + \psi_p r)\,, \qquad p \in [0,1], \tag{2.7}$$

where we have ignored a factor of r_{i} as in equation (2.4). The minimum mortality cost yields the ***group optimum*** coverage level, which we denote p_{g}. The minimum of $C(p)$ on $[0,1]$ may be attained either at a local minimum in $(0,1)$, or at one of the endpoints,

$$C(0) = a(\pi_0 + \psi_0 r)\,, \tag{2.8a}$$
$$C(1) = r\,. \tag{2.8b}$$

To completely specify the cost $C(p)$, we need the probabilities π_p and ψ_p, derived from the epidemiological model (see §2.6 and appendix 2.E), just as for the individual equilibrium. We have found an exact analytical expression for p_{g} in one sub-case (see appendix 2.H) and calculated it numerically in the other cases.

2.6 Epidemiological models

In order to find the group optimum (p_{g}) and individual equilibrium (p_{i}), two key quantities are calculated from the epidemic models: the ***delayer infection probability*** π_p (the probability of a delayer being infected after an outbreak), and the ***delayer vaccination probability*** ψ_p (the probability that a delayer is eventually vaccinated, given an outbreak).

Both π_p and ψ_p depend on the disease dynamics and the post-outbreak vaccination scenario. In the following, we assume that in the absence of post-outbreak vaccination, the standard susceptible-infected-removed (SIR) model is adequate to represent the disease dynamics [87, §4]. The models do not include vital dynamics (births and deaths from all natural causes other than the disease), since the mean serial interval (also called the disease generation time, $t_{\mathrm{ser}} = 22$ days [48, p. 141]) is much smaller than the mean lifetime (~ 80 years in the US [88]). Note that for diseases for which the outbreak time-scale is similar to the mean lifetime, vital dynamics can easily be included in the analysis (*e.g.,* as in [36], where much longer term dynamics were considered).

Let $S(t)$, $I(t)$, $R(t)$ and $V(t)$ be the proportions of susceptible, infected, removed (recovered or dead from smallpox infection) and vaccinated individuals (immune or dead from

vaccination), respectively, at time t. Our basic framework is the $SIRV$ model, described by the differential equations

$$\dot{S} = -\beta SI - \dot{V}\,, \tag{2.9a}$$
$$\dot{I} = \beta SI - \gamma I\,, \tag{2.9b}$$
$$\dot{R} = \gamma I\,, \tag{2.9c}$$
$$\dot{V} \geq 0\,, \tag{2.9d}$$

where $\dot{V}$ must be non-negative as indicated and is defined differently for each of the distinct scenarios of post-outbreak vaccination described in § 2.3.

We assume no one has natural immunity or retains immunity from vaccination decades earlier. This is an approximation, since many living individuals were vaccinated before smallpox was declared eradicated in 1979 [46] and many of those vaccinated individuals are probably still immune (vaccine-derived immunity seems to wane quite slowly and life-long immunity is common [89]). However, smallpox is considered to have been eliminated in the United States as early as 1950, and while routine vaccination continued in some states well after that [46], the proportion of US residents younger than 60 who have been vaccinated is likely very small.

Thus, we assume that the coverage level prior to an outbreak is p, the proportion pre-emptively vaccinated. Consequently, prior to the outbreak, a proportion $1 - p$ of the population is susceptible. When a bioterrorist attack or accidental release takes place (at time $t = 0$), an ***initial attack proportion*** α of the susceptible population is infected. Thus,

$$S(0) = (1-p)(1-\alpha)\,, \tag{2.10a}$$
$$I(0) = (1-p)\alpha\,, \tag{2.10b}$$
$$R(0) = 0\,, \tag{2.10c}$$
$$V(0) = p\,. \tag{2.10d}$$

After an outbreak, the epidemic is over when no one remains infective ($I = 0$). In appendix 2.C, we show rigorously that this is guaranteed to occur, either in finite time or in the limit as $t \to \infty$. In either case we use the subscript ∞ to refer to the time at which the epidemic ends. Thus, S_∞, I_∞, R_∞ and V_∞ refer to the proportions of the population in the susceptible, infective, removed and vaccinated compartments at the end of the epidemic. With this notation, the probabilities of infection and vaccination for delayers are,

respectively,

$$\pi_p = \frac{R_\infty - R(0)}{S_0 + I_0} = \frac{R_\infty}{1-p}\,, \tag{2.11a}$$

$$\begin{aligned}\psi_p &= \frac{V_\infty - V_0}{S_0 + I_0} = \frac{V_\infty - p}{1-p} = 1 - \frac{1 - V_\infty}{1-p} \\ &= 1 - \frac{S_\infty + I_\infty + R_\infty}{1-p} = 1 - \pi_p - \frac{S_\infty}{1-p}\,.\end{aligned} \tag{2.11b}$$

We emphasize that R is the proportion of the population that has been infected (and consequently is either immune or has died), hence $R(0) = 0$ because anyone who is immune at time $t = 0$ is immune from vaccination. Intuitively, there is no endemic equilibrium in these models, because the combination of vaccination and natural spread of disease must eventually cause susceptibles to be so rare that the disease cannot spread (recall that these models neglect vital dynamics).

Lastly, note that π_p is undefined at $p = 1$ (*i.e.,* if everyone pre-emptively vaccinates), as there are no delayers for whom to calculate the probability of being infected. We define π_1 as the limit of the delayer infection probability,

$$\pi_1 = \lim_{p \to 1^-} \pi_p \tag{2.12}$$

i.e., π_1 is the limit of π_p as pre-emptive vaccination approaches full coverage. In appendix 2.D, we show that this limit is equal to the proportion of susceptibles initially infected in an outbreak, *i.e.,* $\pi_1 = \alpha$ for all models considered.

Below we describe (and interpret mechanistically) the various models that we compare, and present some analytical results. In all models, the vaccination rate depends on a vaccination effort parameter, $\phi_{\langle \text{model} \rangle}$, the exact interpretation of which is model-dependent.

2.6.1 Vaccination rate $\propto$ disease prevalence

In this model, vaccination occurs at a rate proportional to disease prevalence (I). A plausible scenario to which such a model would apply is if people respond to media reports on disease prevalence. As a result of increasing disease prevalence, the public might perceive the risk of being infected as higher, and be moved to vaccinate. Consequently,

$$\dot{V} = \phi_{\text{prev}} \operatorname{sign}(S) I\,, \tag{2.13}$$

where

$$\operatorname{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

This model could also represent the case where vaccination rate is proportional to death rate, *i.e.,* people vaccinate in response to media reports on new disease-induced deaths. Since the death rate is proportional to the rate at which the removed compartment, R, grows, which is proportional to I, vaccination rate would be proportional to I as well.

In appendix 2.E.1 we find the final size relations [90, 91, 92] for the model defined by equation (2.13). These are given by

$$S_\infty = \begin{cases} 0 & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ S_\infty^1 & \text{if } p_0 \leq p \leq 1\,. \end{cases} \tag{2.14a}$$

$$R_\infty = \begin{cases} 1 - p - \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right) & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ \frac{\gamma}{\gamma + \phi_{\text{prev}}} \left(1 - p - S_\infty^1\right) & \text{if } p_0 \leq p \leq 1\,. \end{cases} \tag{2.14b}$$

$$V_\infty = \begin{cases} p + \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right) & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ \frac{1}{\gamma + \phi_{\text{prev}}} \left(\phi_{\text{prev}}(1 - S_\infty^1) + \gamma p\right) & \text{if } p_0 \leq p \leq 1\,. \end{cases} \tag{2.14c}$$

with

$$S_\infty^1 = -\frac{1}{\beta}\left(\phi_{\text{prev}} + (\gamma + \phi_{\text{prev}})\, W_0\left(-\frac{\beta S(0) + \phi_{\text{prev}}}{\gamma + \phi_{\text{prev}}} e^{-\frac{\beta(1-p) + \phi_{\text{prev}}}{\gamma + \phi_{\text{prev}}}}\right)\right), \tag{2.15}$$

$$p_m = 1 + \frac{\alpha\phi_{\text{prev}} - \gamma(1-\alpha)}{\beta(1-\alpha)}, \tag{2.16}$$

$$p_0 = 1 + \frac{\phi_{\text{prev}}}{\beta(1-\alpha)} + \frac{\gamma + \phi_{\text{prev}}}{\beta}\, W_k\left(-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma + \phi_{\text{prev}})} e^{-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma + \phi_{\text{prev}})}}\right), \tag{2.17}$$

where $k = 0$ if $p_m < 1$ and $k = -1$ if $p_m \geq 1$. W_0 is the principal branch of Lambert's W function [93, 94], and W_{-1} is its other real branch (see appendix 2.A). p_m is the unique maximum of the function S_∞^1. S_∞^1 has two roots, one at $p = 1$, and the other at p_0, which need not lie in the interval $[0, 1]$ (p_0 is a formal root and need not correspond to a meaningful probability). Note that if $p_m > 1$ then $p_0 > p_m$, and if $p_m < 1$ then $p_0 < p_m$.

If $p_m > 1$ then no delayers will remain susceptible at the end of the epidemic (*i.e.,* all delayers will either be vaccinated or infected), regardless of the initial vaccine coverage level p. Moreover, if $p_m < 1$, but $p_0 < 0$, there are always some delayers who remain susceptible at the end of the epidemic, regardless of the initial coverage level p. If $0 < p_0 < p_m < 1$, then for $p \in [0, p_0]$ there will be no susceptibles left at the end of the epidemic, and for $p \in (p_0, 1)$ there will be some remaining susceptibles. Thus, there is a wide range

of parameter values for which some susceptibles remain at the end of the epidemic; in such cases, $\pi_p + \psi_p < 1$. Numerical evidence and biological intuition suggest that π_p is a decreasing function of p, and we assume that this is the case from here on (this is proven for $p_m \geq 1$ in appendix 2.E.1).

Finally, since the mean infectious period (7 days, see table 2.2 and §2.7.1) is longer than the time required to complete the vaccination program (possibly as short as 3 days [47]), it is interesting to take the limit $\gamma \to 0$ (corresponding to an infinite infectious period) while keeping $\mathcal{R}_0 = \beta/\gamma$ fixed. In this limit, $p_m \to \infty$ so $S_\infty = 0$ (equation (2.14a)), which is in accordance with the assumption—made in [37]—that individuals are ultimately either removed or vaccinated.

We show in appendix 2.G.1 that there is always a unique CSNE, that is, a "best strategy" from the individual perspective. Moreover, an analytical expression for this individual equilibrium can be found if

$$\text{either} \quad \phi_{\text{prev}} \geq \gamma(1-\alpha)/\alpha, \tag{2.18a}$$

$$\text{or} \quad 0 \leq p_0 < p_m \leq 1, \quad \pi_0 > \rho_1 > \pi_{p_0} \quad \text{and} \quad \pi_{p_0} < \rho_2. \tag{2.18b}$$

In addition, we find an analytical formula for the group optimum when $\phi_{\text{prev}} \geq \gamma(1-\alpha)/\alpha$ (see appendix 2.H for details).

2.6.2 Vaccination rate $\propto$ incidence

A vaccination rate proportional to incidence again reflects media-induced vaccination. However, in this model the public reacts to reports of new cases, rather than reports of the total number of sick individuals. Thus,

$$\dot{V} = \phi_{\text{inc}} SI\,. \tag{2.19}$$

In appendix 2.E.2, we show that

$$S_\infty = -\frac{\gamma}{\beta}\, W_0\left(-\frac{\beta(1-p)(1-\alpha)}{\gamma} e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}(1-p)}\right), \tag{2.20a}$$

$$V_\infty = p + \frac{\phi_{\text{inc}}}{\beta+\phi_{\text{inc}}}\big((1-p)(1-\alpha) - S_\infty\big)\,, \tag{2.20b}$$

$$R_\infty = 1 - p - \frac{\phi_{\text{inc}}(1-p)(1-\alpha) + \beta S_\infty}{\beta+\phi_{\text{inc}}}\,. \tag{2.20c}$$

Again, since there are susceptible individuals left at the end of the epidemic, $\pi_p \neq 1-\psi_p$. We show that $\partial_p \pi_p < 0$ (in appendix 2.E.2) and find that there is a unique CSNE, p_{i}, for which an exact formula is derived in appendix 2.G.2.

2.6.3 Vaccination rate $\propto$ proportion still susceptible

In this scenario, susceptible individuals vaccinate at a rate

$$\dot{V} = \phi_{\text{susc}} S \,. \tag{2.21}$$

This is a null model, in the sense that susceptible individuals have a constant probability per unit time of being vaccinated, ϕ_{susc}, independent of the state of the outbreak, as shown in appendix 2.B.1.

We were able to obtain analytical final size relations for this model (see appendix 2.E), but we found the formulae too cumbersome to be useful. Thus, the remainder of our analysis of this model was performed by integrating the differential equations numerically. In our numerical simulations we always find that π_p decreases with p (in appendix 2.G.3, our proof of the existence of a CSNE depends on this being true).

2.6.4 Instantaneous vaccination of a proportion ϕ_{inst} of the population

Some experts believe that the entire United States could be vaccinated in three days [47], which is less than the latent period of smallpox. Consequently, instantaneous vaccination of a proportion ϕ_{inst} of the population remaining susceptible after the outbreak is also a realistic scenario to model. In this case, once vaccination has occurred, the disease simply spreads according to the standard SIR model,

$$\dot{S} = -\beta S I \tag{2.22a}$$
$$\dot{I} = (\beta S - \gamma) I \tag{2.22b}$$
$$\dot{R} = \gamma I \,, \tag{2.22c}$$

with initial conditions given by

$$\begin{aligned} S(0) &= (1-p)(1-\alpha)(1-\phi_{\text{inst}}) \\ I(0) &= (1-p)\alpha \\ R(0) &= 0 \\ V(0) &= p + \phi_{\text{inst}}(1-p)(1-\alpha) \,. \end{aligned}$$

Note that in this scenario we deviate from the convention we use for all the other models, in which $S(0)$ is the initial density of susceptibles prior to the beginning of the post-outbreak vaccination response. Here, $S(0)$ is the density of susceptibles *after* the post-outbreak vaccination response has taken place.

For this scenario, we find (in appendix 2.E.4)

$$S_\infty = -\frac{\gamma}{\beta}\, W_0\left(-\frac{\beta}{\gamma} S(0) e^{-\frac{\beta}{\gamma}(1-V(0))}\right), \tag{2.23a}$$

$$R_\infty = \frac{\gamma}{\beta} \ln \frac{S(0)}{S_\infty}. \tag{2.23b}$$

We also show that π_p is a decreasing function of p, ψ_p is constant and $\pi_p + \psi_p \neq 1$ (see appendix 2.E.4). In addition, we have proved that for this model, there is always a unique CSNE, for which we derive an exact formula in appendix 2.G.4.

2.6.5 Constant rate vaccination

This is the model of Bauch *et al.* [37], in which vaccination occurs at a constant rate ϕ_{const}. Note that in [37] vaccination begins after a response lag t_{lag}, which is the public health services' response time. This lag is taken to be $t_{\text{lag}} = 0$ except in §2.8.5.

For consistency with [37], we included an exposed (but not infective) stage (E), in this model, making it an $SEIRV$ model. This contrasts all the other scenarios, which we have modelled using a simpler $SIRV$ formulation. Our choice of the $SIRV$ framework for the new scenarios is motivated by mathematical tractability and by work subsequent to [37] indicating that $SEIR$ dynamics are captured by an appropriately parameterized SIR model (§2.7.1 below, but see §2.8.5 for an exception).

The model equations for the constant rate vaccination scenario are

$$\dot{S} = -\beta S I - \dot{V} \tag{2.24a}$$

$$\dot{E} = \beta S I - \sigma E \tag{2.24b}$$

$$\dot{I} = \sigma E - \gamma I \tag{2.24c}$$

$$\dot{R} = \gamma I \tag{2.24d}$$

$$\dot{V} = \begin{cases} \phi_{\text{const}} & \text{if } t_{\text{lag}} < t \text{ and } S > 0, \\ 0 & \text{if } t \leq t_{\text{lag}} \text{ or } S \leq 0\,. \end{cases} \tag{2.24e}$$

We have not found a final size relation for this model.

Under the biologically plausible assumption that π_p decreases with p (verified by simulation), [37] have shown the existence of a unique CSNE for this model.

2.7 Parameter estimates, Fair Comparisons of Models and numerical procedures

Because one of the models we investigate includes an exposed class, and the vaccination effort parameter $\phi_{\langle \text{model} \rangle}$ has a different meaning in each scenario we examine, fair comparisons of model results is not completely straightforward. In this section, we consider how the various models can be compared.

2.7.1 SIR vs SEIR

It is well known that similar dynamics are obtained with the standard SIR and SEIR models with identical basic reproductive number, $\mathcal{R}_0$, if the mean infectious period in the SIR model is set equal to the sum of the mean latent and infectious periods in the SEIR model [85, p. 668]. More generally, models can be fairly compared if they have the same mean serial interval [87, §4].

Estimates of the basic reproductive ratio $\mathcal{R}_0$ of smallpox vary in the range $3 \leq \mathcal{R}_0 \leq 10$ [85, 95, 96]. Following [37], we take $\mathcal{R}_0 = 5$. We take the mean serial interval to be $t_{\text{ser}} = 22$ days, as in [48, p. 141] (but note that [37] used $t_{\text{ser}} = 14$ days, and [97] estimated $t_{\text{ser}} = 17.7$ days).

In the constant rate vaccination model, we take the mean latent period to be $1/\sigma = 15$ days [48, p. 141] (based on summing the incubation and prodrom periods, which typically last 12 and 3 days, respectively; see [46, p. 188]). In an SEIR model, the mean serial interval is the sum of the mean latent and infectious periods [87, 98], hence, $1/\gamma = 22 - 15 = 7$ days and $\beta = \gamma \mathcal{R}_0 = \frac{5}{7}$/day. In the SIRV models we take $1/\gamma = t_{\text{ser}}$, while β is modified so that $\mathcal{R}_0 = 5$ (that is, $\beta = \gamma \mathcal{R}_0 = \frac{5}{22}$/day).

2.7.2 Vaccination effort parameter $\phi_{\langle \text{model} \rangle}$

Public health policy changes will affect the vaccination effort parameter $\phi_{\langle \text{model} \rangle}$, where $\langle$model$\rangle$refers to any of "prev", "inc", "susc", "inst" or "const". In order to compare the outcomes of the various vaccination scenarios, for each vaccination model, we find the ***fair comparison value*** of $\phi_{\langle \text{model} \rangle}$, that is, the value of $\phi_{\langle \text{model} \rangle}$ that yields a maximal vaccination rate that is equal to the fixed rate in the constant rate vaccination model of Bauch *et al* [37], $\dot{V} = 0.1$/day (see description under ϕ_{const} below). This allows us to identify, for each scenario, ranges of $\phi_{\langle \text{model} \rangle}$ that can feasibly be attained in reality (*i.e.,* $\phi_{\langle \text{model} \rangle}$ between 0 and the fair comparison value). Our aim is then to compare the different vaccination strategies in terms of vaccine doses used and total expected mortality (we will be interested in the values of these observables at both the individual equilibrium and the group optimum). The fair

comparison values are summarised in table 2.4.

ϕ_{prev} In the prevalence model, $\dot{V} = \phi_{\text{prev}} I$, the vaccination rate is proportional to the prevalence, I, and the vaccination effort parameter ϕ_{prev} is the rate of vaccination *per infected individual*. In appendix 2.F.1, we calculate the maximal vaccination rate as a function of the model parameters, α, β, γ and ϕ_{prev} and p. We find that the maximal vaccination rate for a given initial coverage, p, decreases with the vaccination effort, ϕ_{prev}. We also find that when α, β, and γ are as in Tables 2.1 and 2.2, a maximal vaccination rate of $0.1/\text{day}$ is obtained when $\phi_{\text{prev}} \approx 1582/\text{day}$.

ϕ_{inc} In the incidence model, $\dot{V} = \phi_{\text{inc}} SI$, the vaccination effort parameter ϕ_{inc} is the vaccination rate *per infective per susceptible*. In appendix 2.F.2 we calculate the maximal vaccination rate, as a function of the model parameters, α, β, γ and ϕ_{inc}. We show that the maximal vaccination rate, $\max\{\dot{V} : t \geq 0, p \in [0,1]\}$, is an increasing function of ϕ_{inc}, and that in order to obtain a maximal vaccination rate of $0.1/\text{day}$ or lower, with α, β, and γ as in Tables 2.1 and 2.2, one needs $\phi_{\text{inc}} \approx 5190/\text{day}$.

ϕ_{susc} With $\dot{V} = \phi_{\text{susc}} S$, the vaccination effort parameter ϕ_{susc} is the vaccination rate *per susceptible individual* (alternatively, ϕ_{susc} can be interpreted as the probability per unit time of a delayer being vaccinated; see appendix 2.B.1). In this model, the vaccination rate $\dot{V}$ is always decreasing, since S can only decrease, so $\max\{\dot{V}\} = \phi_{\text{susc}} S(0) = \phi_{\text{susc}}(1-\alpha)(1-p)$ (*cf.* equation (2.10a)). Since the maximal vaccination rate decreases with increasing initial coverage, p, maximal vaccination rate is attained with no pre-emptive vaccination ($p = 0$). Since $S(0) = 1 - \alpha$, the maximal vaccination rate is $\max\{\dot{V}\} = (1-\alpha)\phi_{\text{susc}}$, and a vaccination rate of $0.1/\text{day}$ is attained for $\phi_{\text{susc}} = \frac{0.1}{1-\alpha} \approx 0.1/\text{day}$ (because $\alpha \ll 1$).

ϕ_{inst} For instantaneous vaccination, the vaccination effort parameter ϕ_{inst} is the proportion of susceptibles instantaneously vaccinated when an outbreak occurs. Thus, $\phi_{\text{inst}} \in [0,1]$. The vaccination rate is either 0 (if $\phi_{\text{inst}} = 0$) or effectively infinite (if $0 < \phi_{\text{inst}} \leq 1$, because vaccination occurs all at once). We thus consider the entire range $0 \leq \phi_{\text{inst}} \leq 1$, since there is no value of ϕ_{inst} that results in a vaccination rate of $0.1/\text{day}$.

ϕ_{const} With $\dot{V} = \phi_{\text{const}}$, the vaccination rate is constant, so ϕ_{const} is simply the proportion of the total population that can be vaccinated per unit time. Bauch *et al.* [37] estimated ϕ_{const} for New York City to be

$$\phi_{\text{const}} = (5000 \text{ vaccinators}) \times \left(\frac{200 \text{ people/day}}{\text{vaccinator}} \right) \times \frac{1}{10^7 \text{ people}} = \frac{0.1}{\text{day}} . \qquad (2.25)$$

A rate of $\phi_{\text{const}} = 0.1/\text{day}$ means the entire population can be vaccinated in ten days.

2.7.3 Numerical procedures

When generating figures 2.1 to 2.4, calculations of the following quantities were necessary: the mortality cost, $C(p)$ (equation (2.7)), the group optimum, p_{g} (§ 2.5) and the individual equilibrium, p_{i} (§ 2.4).

To find p_{g}, $C(p)$ was numerically minimized using the `optimize` function in `R` [99]. p_{i} was found by implementing the procedures described in appendix 2.G for the various models, using `R`'s `uniroot` function.

The calculations of both p_{g} and p_{i} depend on π_p and ψ_p, the probabilities of a delayer being infected or vaccinated, respectively (equation (2.11)). For the models in which the vaccination rate is proportional to prevalence or incidence, we used the final size relations reported in §§2.6.1 and 2.6.2, respectively, to calculate π_p and ψ_p. For the remaining models, π_p and ψ_p were obtained by numerically integrating the differential equations using the `deSolve` package [100] in `R` [99].

When generating figure 2.5, for *all* the models π_p and ψ_p were calculated by numerical integration of the differential equations.

2.8 Results and Discussion

2.8.1 Group optimum *vs.* individual equilibrium

Figure 2.1 shows the group optimum p_{g} (red) and individual equilibrium p_{i} (black), as the vaccination effort parameter $\phi_{\langle\text{model}\rangle}$ is varied, for the different models. As expected, the group-optimal coverage is never smaller than the individual equilibrium, and both decrease as $\phi_{\langle\text{model}\rangle}$ is increased. The difference, $p_{\mathrm{i}} - p_{\mathrm{g}}$, tends to grow initially with $\phi_{\langle\text{model}\rangle}$, but eventually decreases to 0 because the coverage at both the group optimum and individual equilibrium always drops to 0 if the vaccination rate parameter $\phi_{\langle\text{model}\rangle}$ is increased sufficiently. It is also evident that the difference between the group-optimal coverage and the individual equilibrium depends strongly on the vaccination model used. In general, this difference is much smaller for the instantaneous and constant rate vaccination models than it is for the other models in which vaccination is affected by the state of the outbreak.

2.8.2 Mortality cost *vs.* vaccination cost

Figure 2.2 shows the mortality cost (proportion of the population that dies, left panel) and the vaccination cost (proportion of the population that is vaccinated by the end of the outbreak, right panel) as functions of the vaccination effort parameter, $\phi_{\langle\text{model}\rangle}$, for the various

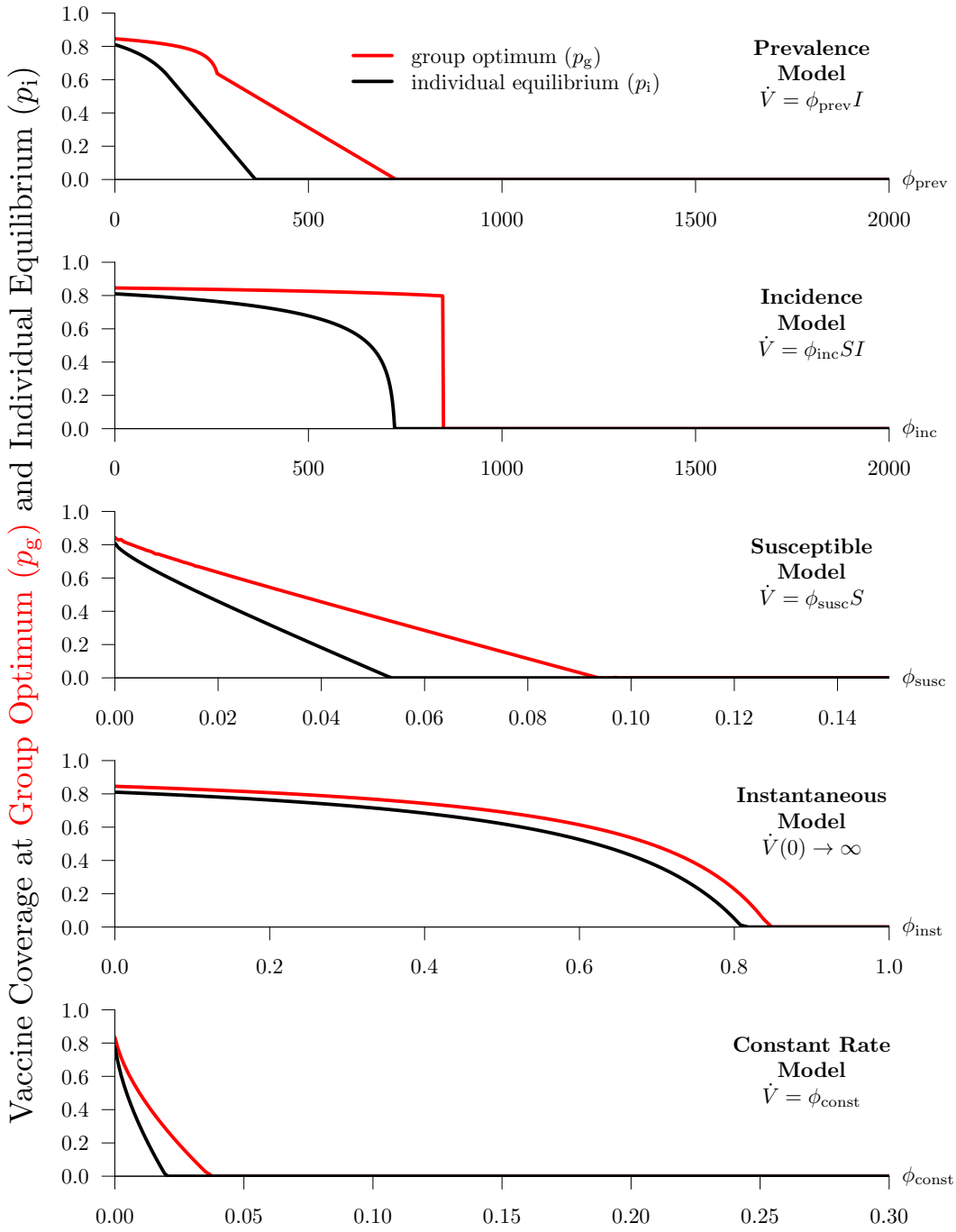

Figure 2.1: Variation of the group optimum p_g (red) and the individual equilibrium p_i (black) with $\phi_{\langle model \rangle}$ (each panel presents results for a different vaccination model). Note the different ranges of $\phi_{\langle model \rangle}$ (on the abscissa) for different models.

post-outbreak response scenarios.

Mortality plateau

The most striking feature of figure 2.2 is the plateau in mortality cost at the individual equilibrium for low values of $\phi_{\langle \text{model} \rangle}$. This plateau can be explained using the Bishop-Cannings theorem [28, 101], which implies that if the individual equilibrium is a mixed strategy then the payoff for vaccinating and delaying must be the same. For low values of $\phi_{\langle \text{model} \rangle}$, the individual equilibrium is mixed ($0 < p_\text{i} < 1$) so the mortality cost associated with vaccinating is the same as for delaying, which is therefore the same as the overall mortality cost. Since the mortality cost for vaccinating is equal to the risk from vaccination (r_v, or r in normalized units; *cf.* equation (2.3), Tables 2.1 and 2.2), the overall mortality cost is constant at r_v (or r in normalized units) as long as the individual equilibrium is mixed. As $\phi_{\langle \text{model} \rangle}$ is increased, the individual equilibrium p_i is decreased (see § 2.8.1). When p_i reaches 0, there is a pure strategy equilibrium (*i.e.,* always delay) so the Bishop-Cannings theorem no longer applies; then the overall mortality is the mortality of delayers, which is $-a[r_\text{i}\pi_p + \psi_p r_\text{v}]$ (see equation (2.1)) and this decreases as $\phi_{\langle \text{model} \rangle}$ is increased (because the epidemic is extinguished faster).

Public health strategy implications of the mortality plateau

There is an important implication of the plateau in mortality that occurs for small $\phi_{\langle \text{model} \rangle}$ if vaccination is voluntary: in order to achieve *any* reduction in overall mortality, the post-outbreak vaccination response must be so strong that no individual would choose to vaccinate pre-emptively ($p_\text{i} = 0$, *i.e.,* the equilibrium is for everyone to delay). Only if the the post-outbreak vaccination response is already sufficiently efficient ($\phi_{\langle \text{model} \rangle}$ is already sufficiently large; figure 2.1) can outbreak size (and hence overall mortality) be reduced by further enhancing the post-outbreak vaccination response (*i.e.,* by increasing $\phi_{\langle \text{model} \rangle}$).

Note that for every model examined here, the right (high effort) edge of the mortality plateau in figure 2.1 occurs for a value of vaccination effort $\phi_{\langle \text{model} \rangle}$ lower than the fair comparison value (see table 2.4). Thus, at the fair comparison values of $\phi_{\langle \text{model} \rangle}$, the individual equilibrium is always to delay vaccination, and mortality can be reduced by increasing vaccination effort, $\phi_{\langle \text{model} \rangle}$.

However, in § 2.8.5, we show that any lag between the start of an outbreak and the beginning of post-outbreak vaccination extends the mortality plateau to higher vaccination efforts, $\phi_{\langle \text{model} \rangle}$, and a long enough lag makes reducing mortality by increasing vaccination effort impossible. We discuss the implications of this for public health strategies further in § 2.8.5.

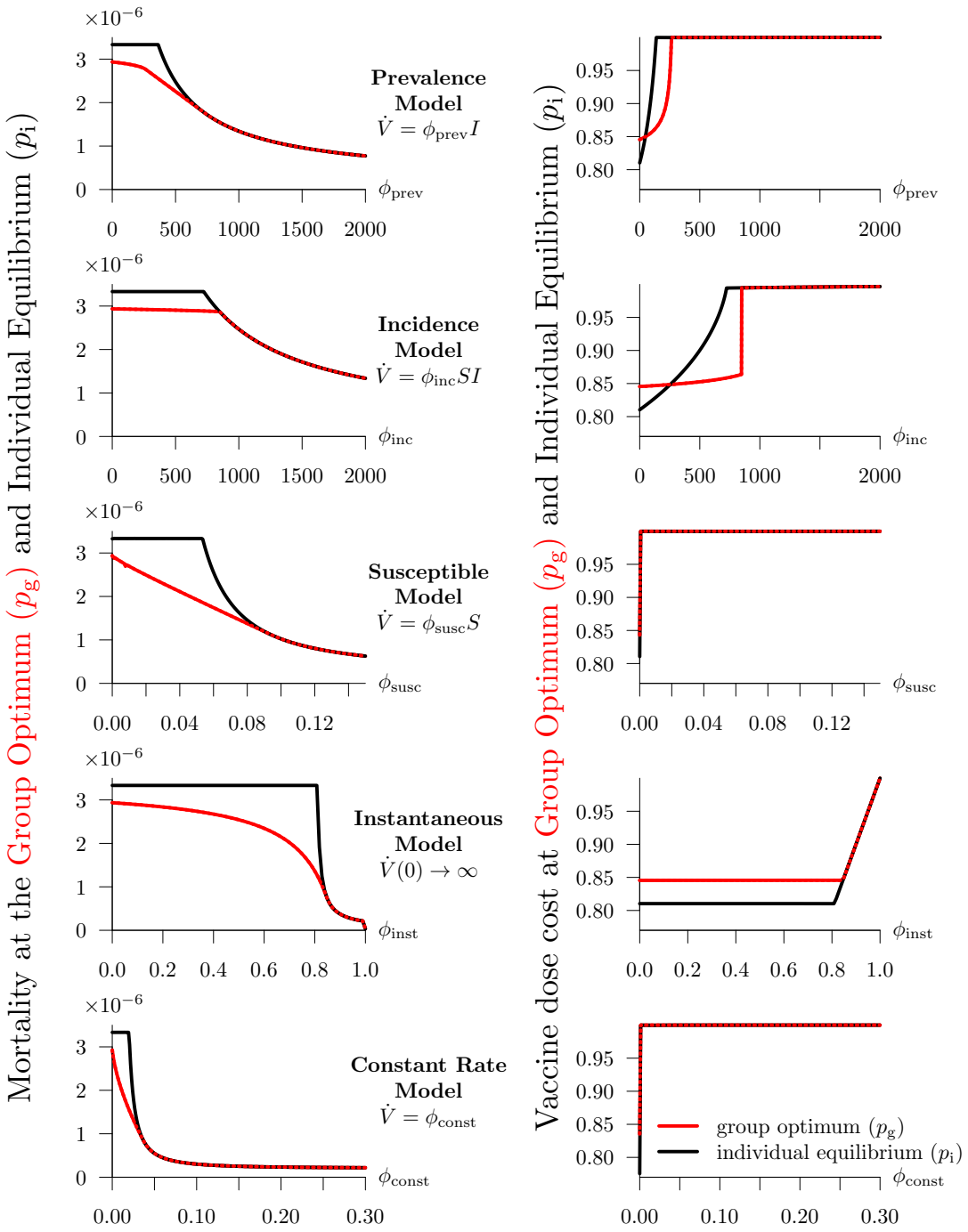

Figure 2.2: Variation of the mortality cost (proportion of the population that dies) and vaccine dose cost (proportion of the population that is vaccinated by the end of the outbreak) as functions of the vaccination effort parameter $\phi_{\langle \text{model} \rangle}$, for different vaccination models. Each row depicts the mortality costs (left panel) and vaccine dose costs (right panel), at the group optimum (red) and individual equilibrium (black) for one model. Note the different ranges of $\phi_{\langle \text{model} \rangle}$ for different models.

Generality of the mortality plateau.

It is important to note that the mortality plateau described earlier is a general phenomenon that applies not only to the post-outbreak vaccination scenarios examined here, but to any reasonable post-outbreak vaccination scenario. More precisely, suppose public health agencies have some measure of control over a vaccination effort parameter, ϕ . Suppose also that $\phi = 0$ corresponds to no possibility of obtaining vaccine post-outbreak, and that the probabilities of a delayer being infected or vaccinated after an outbreak (π_p and ψ_p, respectively) are continuous functions of p and ϕ (for $0 \leq p < 1$ and $\phi \geq 0$). As in §2.4, the costs for delaying and vaccinating individuals are then $a(r_{\text{i}}\pi_p + r_{\text{v}}\psi_p)$ and r_{v}, respectively. Now suppose the following additionally:

1. If there is no possibility of being vaccinated post-outbreak ($\phi = 0$), and no one is vaccinated pre-emptively ($p = 0$), then individuals are at greater risk than if they had vaccinated pre-emptively (*i.e.,* $a(r_{\text{i}}\pi_p + r_{\text{v}}\psi_p)|_{p=0,\phi=0} > r_{\text{v}}$).

2. As the initial coverage approaches 100% ($p \to 1$), the disease does not spread any further than the initial infected cohort ($\pi_p \to \alpha$). Note that as shown in appendix 2.D, this assumption holds for all of the models considered in this paper, and the mathematical argument used to show this is quite general.

3. The risk that a delayer is infected in the initial infection event and then dies, is smaller than the risk of mortality from the vaccine alone ($\alpha r_{\text{i}} < r_{\text{v}}$).

The vaccination game with this post-outbreak vaccination scenario is a population game, and thus must have at least one Nash equilibrium [102, Theorem 2.1.1, p.24]. For low enough vaccination effort ϕ, if coverage p is low, it is more costly to delay than to vaccinate (from the first assumption above). Conversely, if coverage p is high enough, the third assumption above implies that delaying is preferable to vaccinating pre-emptively. It follows that any individual equilibrium that results from the vaccination game is a mixed Nash equilibrium ($0 < p_{\text{i}} < 1$). The preceding argument presented in § 2.8.2 (using the Bishop-Cannings theorem) now implies the existence of a plateau in mortality.

Vaccination cost plateau

The right panels of figure 2.2 also show a plateau for sufficiently *large* vaccination efforts (except for the constant rate vaccination model). Unlike the mortality plateau, this vaccination cost plateau is not rigorously a constant (it changes very slightly as a function of $\phi_{\langle \text{model} \rangle}$), but it is certainly a plateau for all intents and purposes. This plateau occurs because overall vaccination rises with the vaccination effort, $\phi_{\langle \text{model} \rangle}$, and cannot exceed $V_\infty = 1$, so vaccination costs must eventually taper off.

Perceived *vs.* real risks

The general public is likely to overestimate vaccine-induced mortality [103, 104, 105], which would tend to decrease the pre-outbreak vaccine coverage under voluntary vaccination. The game-theoretical framework we employ assumes individuals behave rationally and possess perfect information on which to base their decisions, but it is possible to relax the assumption of perfect information while maintaining that of rationality. Thus, to account for misinformation regarding the dangers of vaccination (possibly as a result of vaccine scares), we can interpret r_{i} and r_{v} as the *perceived* risks of infection and vaccination (rather than the actual risks) to predict the effective level of vaccine coverage prior to an outbreak (note that perceived risks are to be used to predict the individual equilibrium, p_{i}, but not when predicting the group optimum p_{g}, nor when predicting the mortality and vaccination costs at *either* of these coverages). Consequently, public health agencies can potentially reduce mortality by attempting to influence the public's estimate of r (the risk of vaccination relative to infection). For example, risk perception might be influenced by a media campaign aiming to increase the accuracy of the public's perception of vaccine safety and promote pre-emeptive vaccination.

2.8.3 Comparison of relative costs

In figure 2.3 (left panel) we look at the relative mortality cost difference, that is, in units of the cost of optimal mandatory vaccination. Explicitly, we examine how $\frac{C(p_i)-C(p_{\mathrm{g}})}{C(p_{\mathrm{g}})}$ varies with $\phi_{\langle\text{model}\rangle}$ for each model. Similarly, we plot the relative difference in vaccination $\frac{V_\infty(p_i)-V_\infty(p_{\mathrm{g}})}{V_\infty(p_{\mathrm{g}})}$ (figure 2.3, right panel), which is the relative vaccine dose cost difference between voluntary and mandatory vaccination.

Large variation in relative mortality cost. Observe that in figure 2.3 (left panel), the relative mortality cost difference is always non-negative (as expected from the definition of the group optimum as the pre-outbreak coverage for which expected mortality cost is minimal). There is substantial variability among the models in the dependence of the relative mortality cost differences on the vaccination parameter $\phi_{\langle\text{model}\rangle}$. In particular, if vaccination rate is proportional to incidence or prevalence, variation in relative mortality cost is an order of magnitude smaller than if vaccination is instantaneous or at a constant rate. The vaccination scenario that exhibits the largest variation in relative mortality costs is instantaneous vaccination. In this scenario, a voluntary vaccination policy could result in over 150% more deaths than if vaccination were mandatory.

Modest variation in relative vaccine dose cost. There is also substantial variability in the pattern of variation of relative vaccine dose cost as a function of $\phi_{\langle\text{model}\rangle}$ among the dif-

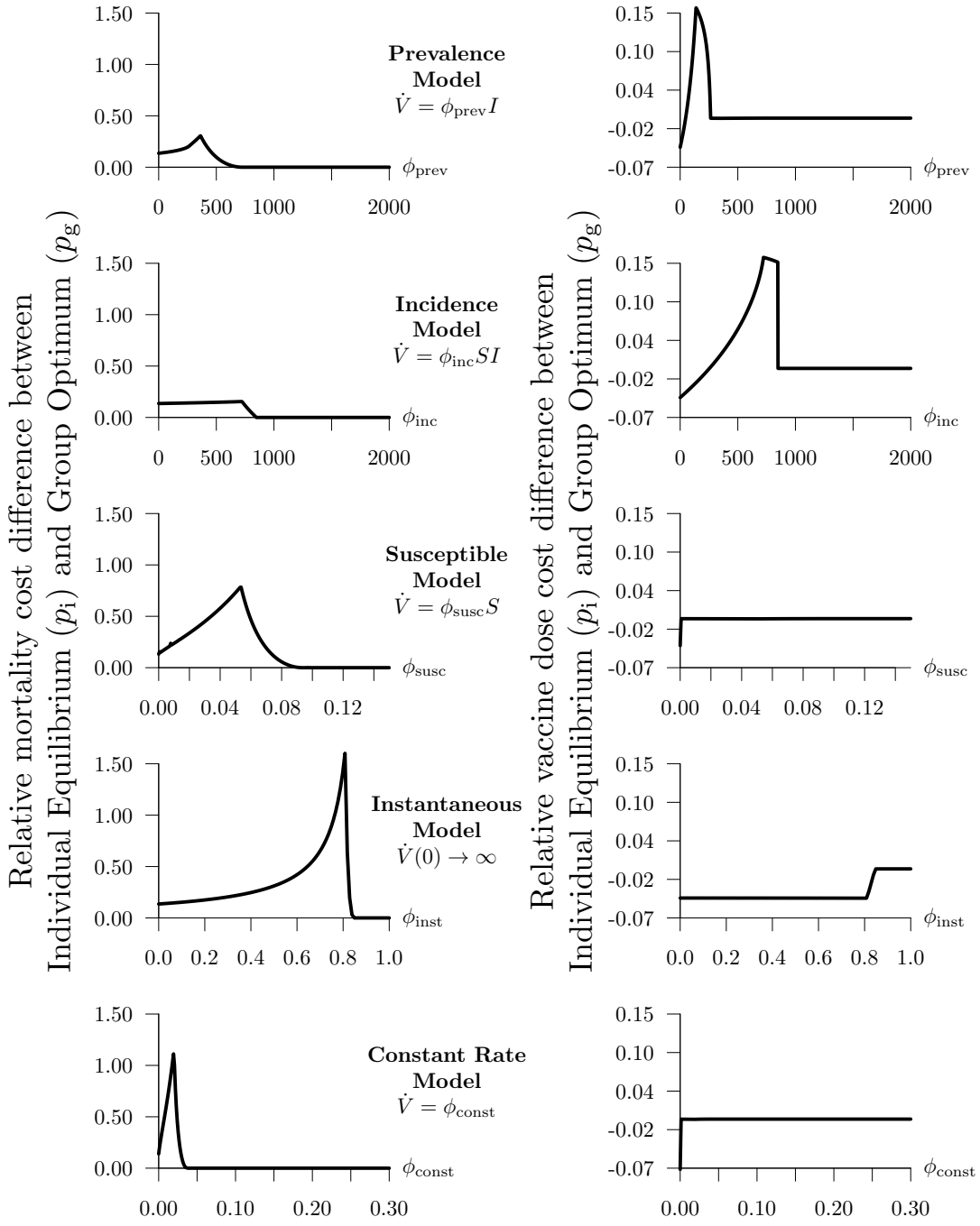

Figure 2.3: Variation of relative difference in mortality and vaccine dose costs with $\phi_{\langle \text{model} \rangle}$, for different vaccination models. Each row depicts the relative mortality cost difference (left panel) and relative vaccine dose cost difference (right panel) for one model. Note the different ranges of $\phi_{\langle \text{model} \rangle}$ for different models.

ferent models. However, for all the models, variation in relative vaccine dose cost as a function of $\phi_{\langle \text{model} \rangle}$ is much less than the corresponding variation in relative mortality cost. The maximum variation in relative vaccine dose cost reaches $\sim 16\%$ for the models in which vaccination is proportional to prevalence or incidence. This relatively large variation can be attributed to low pre-outbreak vaccination coverage (at the individual equilibrium) causing high disease prevalence and incidence; consequently, since vaccination rate is proportional to prevalence or incidence, there is correspondingly high post-outbreak vaccination, which overshoots that which would be required to minimize group mortality. In these two situations, the vaccine dose cost at the individual equilibrium can be greater than at the group optimum. In any case, the relatively small difference in overall vaccine dose costs, both as a function of vaccination effort ($\phi_{\langle \text{model} \rangle}$) and among vaccination scenarios (see right panel of figure 2.2), suggests that vaccine dose cost should likely not be a factor in public health policy decisions.

2.8.4 Vaccine dose cost as a function of mortality cost

Figures 2.2 and 2.3 present mortality costs and vaccine dose costs as functions of vaccination effort for the various models. Because the meaning of the vaccination effort parameter $\phi_{\langle \text{model} \rangle}$ differs among models, it is not straightforward to make meaningful comparisons among the various models (which is why we calculated "fair comparison" values in §2.7). In this section, we display results for the various models, factoring out the vaccination effort parameter. For each model, figure 2.4 shows the vaccine dose cost *as a function of* mortality cost. In health economics terms, this can be considered a ***cost effectiveness analysis***[106].

In figure 2.4, the squares indicate the point in the mortality-cost–vaccination-cost plane where the vaccination effort ($\phi_{\langle \text{model} \rangle}$) is the lowest that we considered. Increasing vaccination effort (while remaining at the individual equilibrium or the group optimum) corresponds to moving away from the square along the plotted curves.

The graphs in figure 2.4 allow us to answer practical questions such as "If we want to ensure that no more than one in every 10 million citizens dies, how many vaccine doses are required in each scenario?" or "If we have a stockpile of vaccine doses sufficient for 30% of the population, what percentage of the population can be expected to die if there is an outbreak in each scenario?" Of course, by construction the graphs do not indicate how much effort ($\phi_{\langle \text{model} \rangle}$) is required to achieve the desired results. We emphasize that—as shown in the previous section—vaccine dose cost at the individual equilibrium or group optimum hardly varies as a function of vaccination effort ($\phi_{\langle \text{model} \rangle}$), so the "practical" questions are not necessarily well-posed (*e.g.,* if we have sufficient vaccine doses for only 30% of the population then neither the individual equilibrium nor the group optimum can ever be achieved). This is true for all the models *with parameters appropriate for smallpox*; for another disease graphs like figure 2.4 could have genuine practical value for public health policy analysis (for example, setting $\mathcal{R}_0 = 1.25$ and keeping all other model parameters as

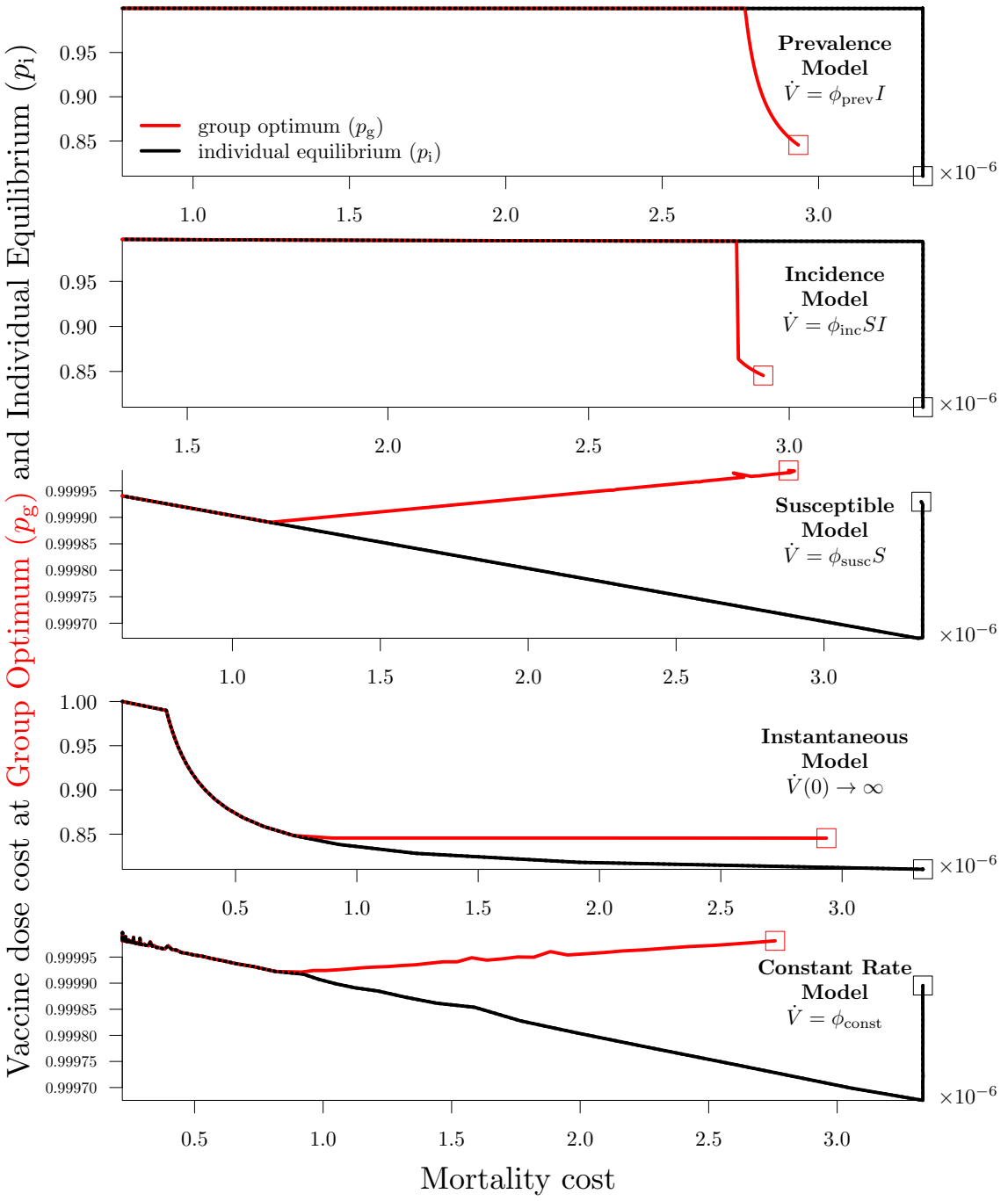

Figure 2.4: Vaccine dose cost *as a function of* mortality cost at the group optimum, p_g (red) and the individual equilibrium, p_i (black), for the different models. Squares represent values at lowest $\phi_{\langle \text{model} \rangle}$ simulated.

in Tables 2.1 and 2.2 causes the vaccination cost to vary between 25% and 99.999%).

In figure 2.4, when the vaccination rate is proportional to either prevalence or incidence, note that as the vaccination effort, $\phi_{\langle \text{model} \rangle}$, increases, two phases of behaviour are apparent for the costs at both p_{i} and p_{g}: first, vaccine dose cost rises but no change in mortality cost is observed (this is caused by the plateau in mortality described in §2.8.2). Then, for all but the instantaneous vaccination model, vaccine dose cost remains virtually constant (note the differences in the scales of the vertical axes among the various panels), but mortality costs decrease.

It is interesting to note that the dependence of vaccine costs on mortality costs at the group optimum varies among the models. For example, when vaccination is proportional to remaining susceptibles, and for the constant rate vaccination model, we see in figure 2.4 that at the group optimum, as mortality cost is decreased, vaccine dose cost decreases at first, but then increases. Thus, in these situations, one can lower both the mortality and the vaccine dose cost at the same time by increasing vaccination effort (in health economics terms, the decision to use higher vaccination effort has negative marginal cost in vaccine doses per life saved). This contrasts the models in which vaccination is instantaneous, or proportional to incidence or prevalence, in which we observe that as mortality cost is decreased, the vaccine dose cost at the group optimum, remains constant and then increases sharply.

Finally, for the instantaneous vaccination model, there is a range of vaccination efforts for which one can reduce mortality without increasing vaccine dose costs at the group optimum. In this parameter range, the increase in vaccine dose cost necessary to decrease mortality at the individual equilibrium is small at first, but grows larger as mortality is decreased.

2.8.5 Effect of vaccination response lag t_{lag}

We have implicitly assumed that in any of the scenarios we have considered the vaccination response will begin as soon as an outbreak is seeded by a bioterrorist attack or accidental release. In contrast, Bauch and co-workers [37] assumed a lag of two weeks between the seeding of an outbreak and the initiation of a vaccination response. In this section, we investigate the effect of a ***response lag*** of t_{lag} days between an outbreak being seeded and the post-outbreak vaccination campaign beginning (so far, we have assumed $t_{\text{lag}} = 0$ days; in [37], $t_{\text{lag}} = 14$ days was assumed).

Intuitively, adding a lag between the beginning of an outbreak and the vaccination response allows the disease to spread unhindered for some time, which increases the probability of delayers being infected, thus decreasing the payoff for delaying. As a result, the individual equilibrium p_{i} increases, which consequently extends the mortality plateau (§2.8.2) to higher values of $\phi_{\langle \text{model} \rangle}$.

The critical lag, t^*_{lag}

For a disease such as smallpox with $\mathcal{R}_0 \sim 5$, the expected final size of an uncontrolled epidemic is greater than 99.9% of the population. If no one is pre-emptively vaccinated, and the response lag after an outbreak is seeded is sufficiently long, almost everyone will have been infected before the response begins, *i.e.,* if t_{lag} is sufficiently long then delayers will almost certainly be infected before they can be vaccinated. Consequently, unless the probability of an outbreak (a) is negligible, delaying will be riskier than vaccinating pre-emptively so the individual equilibrium will not be for everyone to delay: we will certainly have $p_{\text{i}} > 0$. It follows that for response lags longer than some ***critical lag***, t^*_{lag}, mortality cannot be reduced no matter much how much effort is applied in the post-outbreak vaccination response (*i.e.,* the mortality plateau described in § 2.8.2 continues for arbitrarily large values of $\phi_{\langle\text{model}\rangle}$).

A more precise argument allows us to estimate t^*_{lag}. Suppose the initial coverage is $p = 0$ (no pre-emptive vaccination). If the risk of becoming infected and dying is larger than the risk from vaccinating, *i.e.,* $ar_{\text{i}}\pi_0 > r_{\text{v}}$ (or, equivalently, $\pi_0 > r/a$), then delaying will not be the individual equilibrium. For any vaccination scenario, the delayers' probability of being infected by the end of the outbreak (equation (2.11a)) is greater than or equal to their probability of being infected before the vaccination response begins (at time t_{lag}),

$$\pi_0 \geq \Big(I(t_{\text{lag}}) + R(t_{\text{lag}})\Big)\Big|_{p=0} . \tag{2.26}$$

Therefore, if

$$\Big(I(t_{\text{lag}}) + R(t_{\text{lag}})\Big)\Big|_{p=0} > \frac{r}{a} , \tag{2.27}$$

then $\pi_0 > r/a$ and delaying is guaranteed not to be the individual equilibrium. But for *any* post-outbreak vaccination scenario that includes a response lag, when $t < t_{\text{lag}}$ the removed proportion of the population, $R(t)$, follows the standard SIR model solution (with no vaccination). For the SIR model with no vaccination ($p = 0$; a, α and $\mathcal{R}_0$ as in table 2.1), numerical simulation shows that equation (2.27) is satisfied for $t_{\text{lag}} \gtrsim 15.1$ days. Hence, if the public health response lag is 16 days or longer, it is guaranteed that (regardless of the vaccination scenario or corresponding vaccination effort $\phi_{\langle\text{model}\rangle}$), delaying will not be the individual equilibrium.

We emphasize that our estimate of 16 days as an upper bound for t^*_{lag} depends on a number of factors, including:

- The probability of an outbreak (a).
- The proportion of susceptibles infected in the initial outbreak (α).
- The epidemiological model: the estimate is obtained using the SIR model, but adding an exposed class (SEIR) with parameters as in table 2.1 increases the critical lag. Re-

peating the calculation for the SEIR model yielded the upper bound $t^*_{\text{lag}} \leq 26.3$ days. The reason for this difference in critical lags is that when the outbreak is seeded, all individuals initially infected begin their latent period simultaneously, and take on average 15 days to become infectious.

Thus, $t^*_{\text{lag}} < 16$ days should be regarded as a rough estimate at best. Nonetheless, the existence of a critical lag, beyond which it is impossible to reduce mortality by increasing vaccination effort, is an important consideration for public health agencies, in devising contingency plans for post-outbreak vaccination against diseases.

Note, however, that in the case of a bioterrorist attack, an outbreak will probably not be discovered until individuals show symptoms, *i.e.,* until someone's latent period has passed (12 days at a minimum). Taking this delayed detection into account, it follows that in order to avoid extending the mortality plateau to all feasible values of vaccination effort, $\phi_{\langle \text{model} \rangle}$, the response lag from discovery of the epidemic to the beginning of the post-outbreak vaccination response must, in practice, be substantially shorter than 26 days. This is in contrast to an accidental release, where public health authorities might know of the outbreak well before anyone has shown symptoms. In this latter case, because it is more likely that the critical lag has not been exceeded, it is especially important to begin the vaccination response as early as possible in order to reduce mortality.

Lastly, it is important to note that the effect of a response lag on the mortality plateau presupposes that both the vaccination effort and the response lag are known to the public in advance. This limits the applicability of this effect, because in the case of a bioterrorist attack, the response lag likely depends on when an infective first shows symptoms (which introduces a stochastic effect). Further analysis would be needed to determine the effects of a stochastic response lag on individual behaviour, and thus on mortality.

The effective critical lag, $\widetilde{t^*_{\text{lag}}}$

We have seen that if the response lag is longer than the critical lag ($t_{\text{lag}} > t^*_{\text{lag}}$), then no matter how large the vaccination effort ($\phi_{\langle \text{model} \rangle}$), it is impossible to reduce mortality. Of course, in practice, the vaccination effort cannot be arbitrarily large and will be constrained by public health resources. Given a maximum *feasible* vaccination effort, it would be helpful to know how long the response lag can be before the mortality plateau extends to all feasible levels of vaccination effort.

To address this issue, we define the ***effective critical lag***, $\widetilde{t^*_{\text{lag}}}$, to be the minimal response lag, such that the individual equilibrium is no longer to delay (*i.e.,* $p_{\text{i}} > 0$) *given a maximum feasible vaccination effort* $\phi_{\langle model \rangle}$. Thus, the critical lag t^*_{lag} is the limit of $\widetilde{t^*_{\text{lag}}}$ as the maximum feasible vaccination effort becomes arbitrarily large.

In figure 2.5 we plot the effective critical lag $\widetilde{t^*_{\text{lag}}}$, against the vaccination effort $\phi_{\langle \text{model} \rangle}$, for the various models examined in this paper. For the models for which fair comparison

values of $\phi_{\langle \text{model} \rangle}$ are well-defined (see §2.7.2), we used these as estimates for feasible vaccination efforts. However, because the fair comparison level of vaccination effort is a crude estimate for the range of feasible vaccination efforts, in the top panel of figure 2.5 we plot the effective critical lag at values of $\phi_{\langle \text{model} \rangle}$ ranging from 50% to 150% of the fair comparison levels of vaccination efforts for the various models, in increments of 10% of the fair comparison level of $\phi_{\langle \text{model} \rangle}$.

The instantaneous vaccination model was the only model for which a fair comparison value of vaccination effort $\widetilde{\phi_{\text{inst}}}$ could not be defined (see §2.7.2). For this model, we show the effective critical lag $\widetilde{t^*_{\text{lag}}}$ for ϕ_{inst} ranging from 0.8 to 1 (if $\phi_{\text{inst}} < 0.8$ then $\widetilde{t^*_{\text{lag}}} < 1$ day) in the bottom panel of figure 2.5.

We see in figure 2.5 that for some vaccination scenarios, minimizing the response lag t_{lag} is essential: even a short lag extends the mortality plateau to all feasible vaccination effort levels, making it impossible to reduce mortality by increasing effort after the lag. We also note that for some scenarios, a good estimate of the attainable vaccination effort is necessary, because the critical effective lag is very sensitive to the vaccination effort. These two facts further underline the importance of accurately modelling post-outbreak vaccination to inform public health decisions relating to post-outbreak contingency plans. When the response lag is longer than the effective critical lag ($t_{\text{lag}} \geq \widetilde{t^*_{\text{lag}}}$), the only plausible way for public health officials to decrease mortality (while allowing individuals to choose whether or not to vaccinate) is to reduce the relative mortality risk (by decreasing the probability of dying from vaccination, *i.e.,* developing a safer vaccine).

The response lag should be minimized

Based on §§2.8.5 and 2.8.5, reducing the response lag lowers expected mortality and makes it easier to decrease mortality further:

- For vaccination efforts higher than the end of the mortality plateau, everyone will choose to delay vaccination (§2.8.2). From the discussion leading up to equation (2.26), it follows that even in the best-case scenario where the epidemic is stopped immediately at $t = t_{\text{lag}}$, mortality will be no less than

$$r_{\text{i}} \left(I(t_{\text{lag}}) + R(t_{\text{lag}}) \right) \Big|_{p=0} , \tag{2.28}$$

  which increases with the response time t_{lag}. Thus, increasing the response lag increases the lowest attainable mortality (even if vaccination effort can be increased without bound).

- Increasing the response lag increases the vaccination effort at the end of the mortality plateau (*i.e.,* the minimal vaccination effort beyond which increasing vaccination effort decreases mortality). Thus, longer response lags make it harder to achieve a

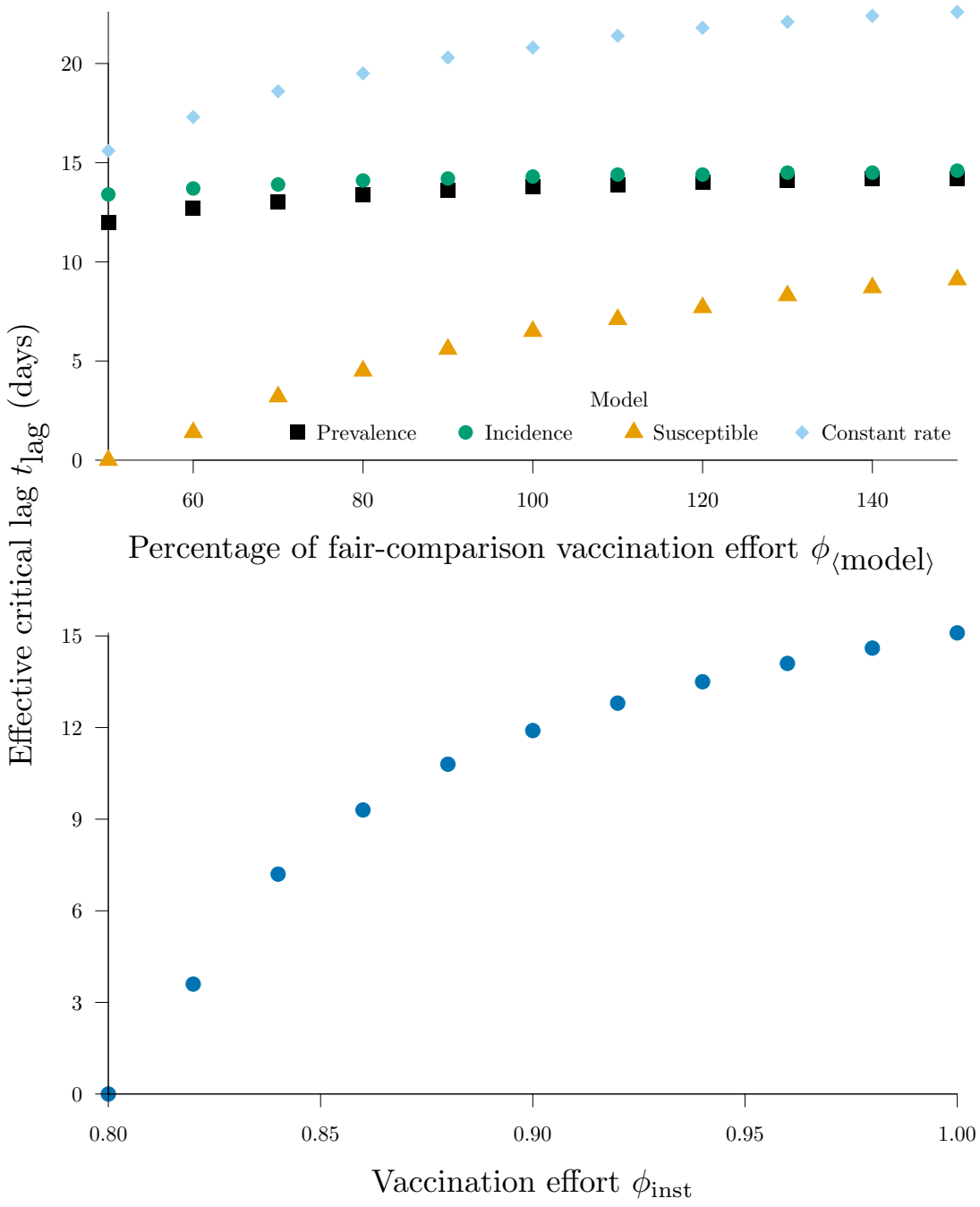

Figure 2.5: Variation of the effective critical lag $\widetilde{t^*_{\text{lag}}}$ (§2.8.5) with vaccination effort $\phi_{\langle \text{model} \rangle}$. Top: effective critical lag *versus* percentage of the fair comparison vaccination effort when the fair comparison vaccination effort is defined. Bottom: effective critical lag *versus* vaccination effort ϕ_{inst} when vaccination is instantaneous.

decrease in mortality.

However, note that if a response time lower than the effective critical lag ($t_{\text{lag}} < \widetilde{t^*_{\text{lag}}}$) cannot be achieved, neither increasing the vaccination effort $\phi_{\langle \text{model} \rangle}$, nor decreasing the response time t_{lag}, can decrease mortality.

2.9 Conclusions

We have analyzed five distinct scenarios (§ 2.3) associated with a potential smallpox outbreak triggered by a bioterrorist attack or accidental release. The scenarios differ in the factors that influence individuals' perception of risk and how a post-outbreak vaccination response plays out. We examined these scenarios both with and without an assumed lag between an outbreak starting and a public health response being initiated. Our work generalizes the analysis of Bauch and co-workers [37] who investigated a single scenario with a response lag of 14 days.

As in [37], we considered separately group interest (optimal strategies for minimizing overall mortality) and self-interest (stable strategies for individual choices with respect to pre-emptive vaccination). From each perspective, we obtained the (imposed or expected) pre-emptive vaccination coverage (p) for each scenario (the group optimum p_{g} in the case of group interest and the individual equilibrium p_{i} in the case of self-interest) (figure 2.1).

Our principal conclusions are the following.

1. For a given level of post-outbreak vaccination effort, the group optimum pre-emptive coverage is always greater than the individual equilibrium ($p_{\text{g}} > p_{\text{i}}$; figure 2.1) and the expected total mortality is always less if public health authorities impose the group optimum rather than letting individuals make their own vaccination decisions (figure 2.2, left column). If no outbreak occurs, then some people will die unnecessarily from pre-emptive vaccination. Given the difficulty of estimating the probability of an attack or accidental release, it would be hard for governments to justify an imposed pre-emptive vaccination policy for a disease like smallpox for which the vaccine can cause death.

2. The number of vaccine doses required at the group optimum and individual equilibrium does not vary substantially as a function of vaccination effort (*e.g.,* speed of vaccine distribution post-outbreak) for any of the scenarios (figure 2.2, right column). Consequently, the economic cost of vaccine production is not likely to play a significant role in policy decisions.

3. Total expected mortality as a function of vaccination effort depends strongly on which scenario is considered (figure 2.2, left column). Some vaccination scenarios are affected by the public reaction to media reports on the epidemic's progress, while some

(the instantaneous and constant rate vaccination scenarios) are under direct control of public health authorities. To assist public health authorities preparing for potential outbreaks, further research is needed to determine which factors have the greatest influence on individuals' perception of risk and which post-outbreak vaccination strategies are most feasible.

4. For any realistic vaccination scenario, there is a range of vaccination effort levels in which increasing vaccination effort does not reduce overall mortality. In this mortality plateau, increasing vaccination effort leads only to fewer people vaccinating preemptively, until the individual equilibrium becomes to delay vaccination (at which point it is possible to reduce mortality by increasing the vaccination effort). Thus, under voluntary vaccination, in order for public health authorities to expect to reduce mortality by increasing vaccination effort post-outbreak, their planned post-outbreak vaccination response must be so efficient that no-one would choose to vaccinate preemptively ($p_{\mathrm{i}} = 0$).

5. Any lag between the beginning of an outbreak and the post-outbreak vaccination response makes it harder for higher vaccination effort levels to make a difference to overall mortality, and a large enough lag will make it impossible to reduce mortality *regardless* of the level of vaccination effort. Given a maximum feasible vaccination effort level, there is an effective critical lag, beyond which it is impossible to reduce mortality by increasing vaccination effort. The dependence on the post-outbreak vaccination scenario, of both the effective critical lag at feasible levels of vaccination effort, and the effect of changes in vaccination effort on the effective critical lag, further highlights the importance of researching realistic post-outbreak vaccination responses.

It is not possible to know with certainty how governments and health agencies will react, or how individuals will behave, in the event of an outbreak. However, the above conclusions are based on our analysis of five distinct post-outbreak scenarios (and some model features that are much more generic), so it seems likely that our conclusions would remain valid if further plausible scenarios were considered.

Acknowledgements

Some preliminary work on the problem addressed in this paper was carried out by Adelicia Yu as part of her undergraduate summer research project in 2004. We are grateful to Sigal Balshine, Paul Higgs, Rufus Johnstone, and two anonymous referees for valuable comments. We were supported by NSERC (DE) and the Trillium Foundation (CM).

Tables

Estimated Parameters

Quantity	Interpretation	Value	Source
r_{v}	Mortality risk from vaccination (probability)	10^{-6}	[37]
r_{i}	Mortality risk from infection (probability)	0.3	[37]
$\mathcal{R}_0$	Basic reproductive ratio	5	[85, 95, 96]
t_{ser}	Mean serial interval	22 days	[48, p. 141]
$1/\sigma$	Mean latent period (SEIRV)	15 days	[46, p. 188], [37]
$\phi_{\langle \mathrm{model} \rangle}$	Vaccination effort parameter (exact interpretation depends on model)	See table 2.4	
t_{lag}	Response lag before initiation of post-outbreak vaccination	0 days, except in §2.8.5.	
a	Probability of attack or accidental release per lifetime	0.01	[37]
α	Proportion of susceptibles initially infected in an outbreak	$\frac{5000}{290\times 10^6} \simeq 1.72 \times 10^{-5}$	[37]

Table 2.1: Summary of the fundamental (*i.e.,* not derived) numerical parameters in our analysis, together with estimated values. Note that in [37] the probability of an outbreak was denoted r rather than a. Here, we use r for the relative risk, as in [36]. The proportion of the population infected initially by a bioterrorist attack or accidental release, α, corresponds to infection of 5000 individuals in a population of 290 million (after [37]).

Derived Parameters

Quantity	Interpretation	Value
$r = r_{\mathrm{v}}/r_{\mathrm{i}}$	Relative risk (from being vaccinated compared with natural infection)	$10^{-6}/0.3 \simeq 3.33\times10^{-6}$
$1/\gamma$	Mean time infectious (SIRV)	$t_{\mathrm{ser}} = 22$ days
$1/\gamma$	Mean time infectious (SEIRV)	$t_{\mathrm{ser}} - (1/\sigma) = 7$ days
β	Transmission rate	$\gamma\mathcal{R}_0$
π_p	Probability that an un-vaccinated individual will eventually be infected if the vaccine coverage level in the population is p	Derived from epidemic model in §2.6
ψ_p	Probability of an individual un-vaccinated at the beginning of the epidemic becoming vaccinated, given vaccine coverage level p	Derived from epidemic model in §2.6

Table 2.2: Summary of derived parameters.

Other Notation

Quantity	Interpretation
P	Probability that an individual chooses to vaccinate pre-emptively (this defines the individual's strategy)
p	Pre-outbreak vaccine coverage (proportion of the population vaccinated pre-emptively)
p_{g}	The group optimum, *i.e.,* the proportion of the population vaccinated pre-emptively which minimizes mortality
p_{i}	The individual equilibrium, *i.e.,* the level of pre-outbreak vaccine coverage which is the unique Nash Equilibrium, as described in §2.4
$C(p)$	The mortality cost, *i.e.,* the proportion of the population that is expected to die, given pre-emptive vaccine coverage p
t^*_{lag}	The critical lag, *i.e.,* the response lag beyond which mortaliy is independent of vaccination effort (see §2.8.5)
$\widetilde{t^*_{\mathrm{lag}}}$	The effective critical lag, *i.e.,* the response lag beyond which mortaliy is identical for all feasible values of vaccination effort (see §2.8.5)

Table 2.3: Summary of other notation.

Vaccination Effort, $\phi_{\langle \mathbf{model} \rangle}$

Model	"Fair Comparison" value	Value at end of mortality plateau
ϕ_{prev}	$1582/\mathrm{day}$	$571/\mathrm{day}$
ϕ_{inc}	$5190/\mathrm{day}$	$1137/\mathrm{day}$
ϕ_{susc}	$0.1/\mathrm{day}$	$0.08/\mathrm{day}$
ϕ_{inst}	—	$0.82/\mathrm{day}$
ϕ_{const}	$0.1/\mathrm{day}$	$0.015/\mathrm{day}$

Table 2.4: Summary of notable levels of the vaccination effort parameter, $\phi_{\langle \mathrm{model} \rangle}$, for the different models. The first column contains "Fair Comparison" values for the vaccination effort parameters of the various models, as calculated in § 2.7.2. In our simulations, we allowed $\phi_{\langle \mathrm{model} \rangle}$ to range between 0 and values generally above the "Fair Comparison" values (except for ϕ_{inst}, for which we used the entire possible range of $[0, 1]$). The second column contains the minimal values of the vaccination effort parameter ($\phi_{\langle \mathrm{model} \rangle}$) for which the individual equilibrium is to delay (that is, $\phi_{\langle \mathrm{model} \rangle}$ at the end of the mortality plateau, see §2.8.2).

Appendix

2.A Lambert W function

The standard final size relation, which can be derived from the SIR model [90] and many other epidemic models [91], is

$$Z = 1 - e^{-\mathcal{R}_0 Z}\,. \tag{2.29}$$

Here, Z is the final size ($Z = 1 - S_\infty$) and $\mathcal{R}_0$ is the basic reproduction number. Z can be expressed explicitly as a function of $\mathcal{R}_0$ [91, 92],

$$Z(\mathcal{R}_0) = 1 + \frac{1}{\mathcal{R}_0}\, W[-\mathcal{R}_0\, e^{-\mathcal{R}_0}]\,, \tag{2.30}$$

where the Lambert W function [93, 94] is the inverse function of

$$f(W) = W\, e^W\,. \tag{2.31}$$

Use of the Lambert W function is critical for our derivations of final size formulae for models we consider here. $W(x)$ is real for $x \geq -1/e \simeq -0.368$ and is two-valued for $-1/e < x < 0$. The upper "principal" branch, for which $W(x) \geq -1$, is denoted W_0 and the lower branch is denoted W_{-1}. See figure 2.A.1.

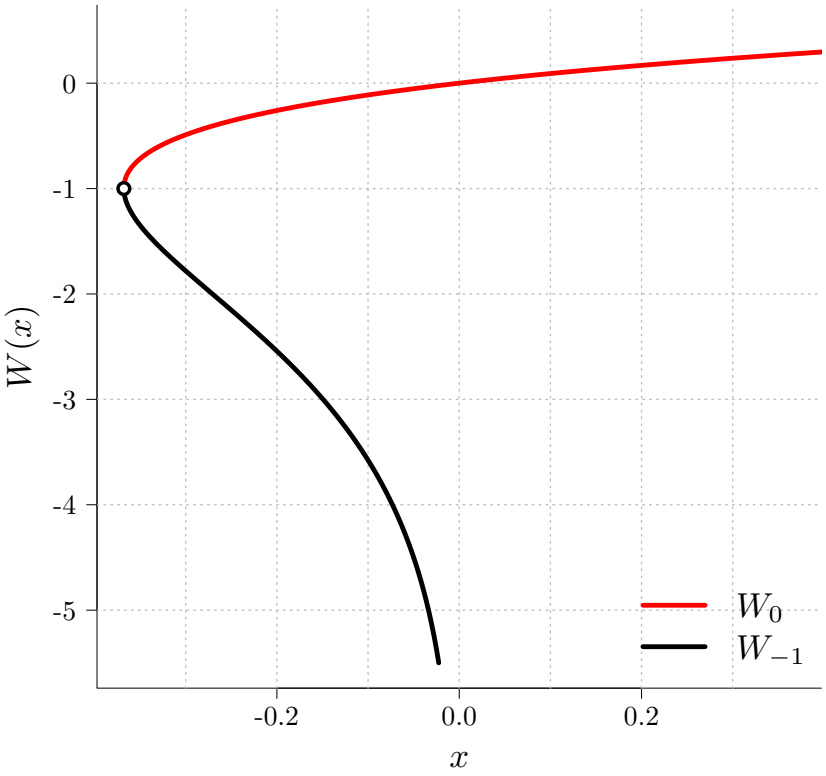

Figure 2.A.1: The Lambert W function, showing the principal branch W_0 and the secondary branch W_{-1}.

2.B Interpretation of vaccination effort parameters

The vaccination effort parameters are explained § 2.6. "Fair comparison" values for these parameters are derived in § 2.7.2 and listed in table 2.4.

2.B.1 ϕ_{susc}

In § 2.7.2, we commented that ϕ_{susc} can be considered to be the probability per unit time of a delayer being vaccinated. To see this, note that the probability (p_{vacc}) of a susceptible delayer being vaccinated in the time interval $[t, t+\Delta t]$ is the ratio of the number of susceptibles vaccinated in that time interval, $V(t+\Delta t) - V(t)$, to the number of susceptibles present at the beginning of that time interval, $S(t)$. Thus,

$$p_{\text{vacc}} = \frac{V(t+\Delta t) - V(t)}{S(t)} = \frac{V(t+\Delta t) - V(t)}{\Delta t} \frac{1}{S(t)} \Delta t \,.$$

Since $\lim_{\Delta t \to 0} \frac{V(t+\Delta t)-V(t)}{\Delta t} = \dot{V} = \phi_{\text{susc}} S$, for small Δt, we have $p_{\text{vacc}} \approx \phi_{\text{susc}} \Delta t$. Thus, ϕ_{susc} is the (constant) probability per unit time of a delayer being vaccinated.

2.C Convergence to disease–free equilibrium

In this appendix, we show that for all models considered, the epidemic must eventually die out (*i.e.,* the system converges to a disease-free equilibrium).

Consider the $SIRV$ model given by the differential equations

$$\dot{S} = -\beta SI - \dot{V} \,, \tag{2.32a}$$
$$\dot{I} = \beta SI - \gamma I \,, \tag{2.32b}$$
$$\dot{R} = \gamma I \,, \tag{2.32c}$$
$$\dot{V} = f(t, S, I, R, V) \,, \tag{2.32d}$$

where f is continuously differentiable and satisfies $f(t, S, I, R, V) \geq 0$, $f(t, 0, I, R, V) = 0$ whenever $t \geq 0$, $S \geq 0$, $I \geq 0$,$R \geq 0$, $V \geq 0$. From the fundamental existence and uniqueness theorem [107], there is a unique solution to equation (2.32) for any non-negative initial conditions $S(0) \geq 0$, $I(0) \geq 0$, $R(0) \geq 0$ and $V(0) \geq 0$. Suppose also that $S(0) + I(0) + R(0) + V(0) = 1$.

First, note that $S(t) \geq 0$ for all $t \geq 0$. To see this, suppose in order to derive a contradiction that $S(T) < 0$ for some $T > 0$. Then, since $S(t)$ is continuous, there must be

some time $0 \leq \tau < T$ such that $S(\tau) = 0$. But because $\dot{S}|_{S=0} = 0$ it follows that $S(t) = 0$ for all $t \geq \tau$, which contradicts $S(T) < 0$.

Similarly, it follows that $I(t) \geq 0$. Consequently, $R(t)$ is nondecreasing in t, and in particular, $R(t) \geq 0$. Lastly, $\dot{V} \geq 0$ so $V(t) \geq 0$ as well.

Now, $\frac{\mathrm{d}}{\mathrm{d}t}\Big(S(t) + I(t) + R(t) + V(t)\Big) = 0$ for all t, so $S(0) + I(0) + R(0) + V(0) = 1$ implies $S(t) + I(t) + R(t) + V(t) = 1$ for all t. Consequently, $S(t)$, $I(t)$, $R(t)$ and $V(t)$ each lie in the interval $[0, 1]$ for all time.

In addition to being bounded, $S(t)$, $R(t)$ and $V(t)$ are monotonic (their time derivatives are non-positive) and therefore have a limit as $t \to \infty$. It follows that $I(t)$ also has such a limit ($I = 1 - S - V - R$). To see that this limit is 0, suppose instead that $I_\infty = \lim_{t\to\infty} I(t) > 0$. Then, $\lim_{t\to\infty} \dot{R} = \gamma I_\infty$. Thus, there exists a time t_* such that $\dot{R}(t) > \gamma I_\infty/2$ for all $t_* < t$. This implies that the proportion in the recovered class increases at least linearly , and must eventually hit $R = 1$ (no later than at time $t_* + 2/(\gamma I_\infty)$) and be greater than 1 thereafter. However, this contradicts the fact that the proportion of the population in any class cannot exceed 1. Thus $I_\infty = \lim_{t\to\infty} I(t) = 0$.

A similar argument can be applied to the constant rate $SEIRV$ model by noting that

1. If $S(0) \geq 0$, $E(0) \geq 0$, $I(0) \geq 0$, $R(0) \geq 0$ and $V(0) \geq 0$, then S, E, I, R and V remain non-negative for all time.
2. If $S(0) + E(0) + I(0) + R(0) + V(0) = 1$, then $S + E + I + R + V = 1$ for all time.
3. $S(t) \to 0$ in finite time in this model. To see this, suppose in order to derive a contradiction, that $S(t) > 0$ for all time $t \geq 0$. Then, for all $t > t_{\text{lag}}$, $\dot{S} = \phi_{\text{const}}$, and thus $S(t) \geq \phi_{\text{const}}(t - t_{\text{lag}})$. Consequently, $S(t) > 1$ for all $t > t_{\text{lag}} + 1/\phi_{\text{const}}$, contradicting the fact that $S(t) \leq 1$ for all t.
4. Since $S(t) \to 0$ in finite time in this model, it follows that after a finite time $\dot{E} = -\sigma E$, implying that $\lim_{t\to\infty} E = 0$. Since R and V are monotonic and thus have a limit as $t \to \infty$, it follows that $\lim_{t\to\infty} I = \lim_{t\to\infty} 1 - S - E - R - V$ exists as well, and one can continue as before.

Lastly, a corollary of this convergence to the disease-free equilibrium is that $S_\infty = \lim_{t\to\infty} S(t)$ is well defined and

$$S_\infty < \frac{\gamma}{\beta} = \frac{1}{\mathcal{R}_0}\,. \tag{2.33}$$

To see this, observe that since S is monotonic and bounded, it must converge to some finite limit within $[0, 1]$, so S_∞ is well defined. Next, note that $I(0) = \alpha(1 - p) > 0$ and $\lim_{t\to\infty} I(t) = I_\infty = 0$, so there is some time t_* at which $\dot{I}(t_*) < 0$ and so $S(t_*) < \gamma/\beta$. But $\dot{S} \leq 0$ and so for any $t_* < t$, we have $S(t) < \gamma/\beta$. Thus, equation (2.33) follows because of the monotonicity of S.

2.D Calculation of π_1

In § 2.6, we stated that as pre-emptive vaccination approaches full coverage ($p \to 1$), the probability of a delayer being infected (π_p) approaches the proportion of susceptibles initially infected in an outbreak (α). Recalling the definition of π_1 in equation (2.12), the claim is that for all the models considered

$$\pi_1 = \lim_{p \to 1^-} \pi_p = \alpha \,. \tag{2.34}$$

To verify equation (2.34), first consider the SIRV models defined in equation (2.9). The proportion of individuals who are eventually removed (R_∞) must be greater than the number initially infected (equation (2.10b)), so

$$R_\infty \geq I(0) = \alpha(1-p) \,. \tag{2.35}$$

Thus from equation (2.11a), we have $\pi_p \geq \alpha$ for any $p \in [0, 1)$. It follows that $\pi_1 \geq \alpha$ if the limit exists.

We now show that $\pi_1 \leq \alpha$. The basic reproduction number is $\mathcal{R}_0 = \beta/\gamma$ and the effective reproduction number when the outbreak begins is (equation (2.10a))

$$\mathcal{R}_{\text{eff}}(0) = \mathcal{R}_0 S(0) = \mathcal{R}_0(1-p)(1-\alpha) \,. \tag{2.36}$$

Thus, if $p > 1 - 1/\mathcal{R}_0$ then $\mathcal{R}_{\text{eff}}(0) < 1$, and

$$\lim_{p \to 1^-} \mathcal{R}_{\text{eff}}(0) = 0 \,. \tag{2.37}$$

To prove that $\pi_1 \leq \alpha$, we will show that

$$\pi_p \leq \frac{\alpha}{1 - \mathcal{R}_{\text{eff}}(0)} \,, \qquad \text{for all } p > 1 - \frac{1}{\mathcal{R}_0} \,. \tag{2.38}$$

Equations (2.9b) and (2.32c) imply that $\dot{R} = \gamma I = \beta SI - \dot{I}$. Hence,

$$R_\infty = \int_0^\infty \dot{R} \, \mathrm{d}t = \int_0^\infty \left(\beta SI - \dot{I}\right) \mathrm{d}t \,.$$

But $S(t)$ decreases monotonically with t, and $I_\infty = 0$ (appendix 2.C), so

$$R_\infty \leq \beta S(0) \int_0^\infty I \, \mathrm{d}t + I(0) = \mathcal{R}_{\text{eff}}(0) \int_0^\infty \gamma I \, \mathrm{d}t + I(0) = \mathcal{R}_{\text{eff}}(0) R_\infty + \alpha(1-p) \,. \tag{2.39}$$

Thus,

$$\left(1 - \mathcal{R}_{\text{eff}}(0)\right) R_\infty \leq \alpha(1-p) \,, \tag{2.40}$$

and consequently, for any $p > 1 - 1/\mathcal{R}_0$,

$$\pi_p = \frac{\mathcal{R}_\infty}{1-p} \leq \frac{\alpha}{1 - \mathcal{R}_{\text{eff}}(0)} \,. \tag{2.41}$$

In the limit $p \to 1^-$, equation (2.37) implies $\pi_1 \leq \alpha$, as required.

To see that $\pi_1 = \alpha$ for the constant rate vaccination (SEIRV) model (equation (2.24)), we need only note that in this case,

$$R_\infty = \int_0^\infty \gamma I \, \mathrm{d}t = \int_0^\infty (\beta SI - \dot{I} - \dot{E}) \, \mathrm{d}t \leq \mathcal{R}_0 S(0) \int_0^\infty \gamma I \, \mathrm{d}t + E(0) \,, \tag{2.42}$$

where $E(0) = \alpha(1-p)$. Thus, Inequality (2.39) holds for the SEIRV model as well, and the remainder of the proof that $\pi_1 = \alpha$ is identical to the argument for SIRV models.

2.E Final size relations, π_p and ψ_p

2.E.1 Vaccination rate $\propto$ disease prevalence

Final size relations

A naïve model in which vaccination is proportional to prevalence is

$$\dot{S} = -\beta SI - \phi_{\text{prev}} I \tag{2.43a}$$
$$\dot{I} = \beta SI - \gamma I \tag{2.43b}$$
$$\dot{R} = \gamma I \tag{2.43c}$$
$$\dot{V} = \phi_{\text{prev}} I \,. \tag{2.43d}$$

However, equation (2.43a) is not biologically sensible, since if $S = 0$ and $I > 0$ it follows that $\dot{S} < 0$ and so S attains negative values. Thus, a more realistic model is obtained by replacing the vaccination rate $\phi_{\text{prev}} I$ with $\phi_{\text{prev}} f(S) I$, where f is a nondecreasing and smooth "cutoff function" such that $f(S) = 1$ except for $0 \leq S < \delta$, and $f(0) = 0$. Thus, equation (2.43a) is replaced by

$$\dot{S} = -\beta SI - \phi_{\text{prev}} f(S) I \,. \tag{2.43a$'$}$$

For convenience, we also choose f to be an odd function, *i.e.*, $f(-S) = -f(S)$ (however, since negative values of S are not biologically feasible and are unattainable by this model if $S(0) \geq 0$, this has no effect on the dynamics of the model for biologically sensible initial conditions).

As $\delta \to 0$, $\dot{V}$ approaches $\phi_{\text{prev}} \operatorname{sign}(S) I$ and equation (2.43a′) approaches

$$\dot{S} = -\beta S I - \phi_{\text{prev}} \operatorname{sign}(S) I \,,$$

where

$$\operatorname{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Thus, a more biologically sensible model where vaccination is proportional to prevalence is:

$$\dot{S} = -\beta S I - \phi_{\text{prev}} \operatorname{sign}(S) I \tag{2.44a}$$
$$\dot{I} = \beta S I - \gamma I \tag{2.44b}$$
$$\dot{R} = \gamma I \tag{2.44c}$$
$$\dot{V} = \phi_{\text{prev}} \operatorname{sign}(S) I \,. \tag{2.44d}$$

In the interior of the biologically meaningful domain,

$$\Delta = \{(S, I, R, V) | S \geq 0 \,, I \geq 0 \,, R \geq 0 \,, V = 1 - S + I + R\} \,, \tag{2.45}$$

the phase portrait for this model is similar to that of the original model and the dynamics change only as the hyper-plane $S = 0$ is reached. For this reason, we analyze the original model (equations (2.43)) and make the necessary corrections to account for equation (2.44d) afterwards. We denote state-variable solutions to the original model (equations (2.43)) with a superscript 1, as in S^1, *etc.*

From appendix 2.C, we know that solutions of equations (2.43) converge to a disease-free equilibrium. Thus, we wish to obtain final size relations for this model. We proceed as follows:

From

$$\frac{\mathrm{d}R^1}{\mathrm{d}S^1} = -\frac{\gamma}{\beta S^1 + \phi_{\text{prev}}} \,, \tag{2.46}$$

we have

$$R^1_\infty = \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\beta S^1_\infty + \phi_{\text{prev}}}\right) , \tag{2.47}$$

where a subscript ∞ indicates the value of that variable at the end of the epidemic (recall

that S(0) also depends on p). S^1_∞ is obtained by a similar trick.

$$\frac{\mathrm{d}I^1}{\mathrm{d}S^1} = -\frac{\beta S^1 - \gamma}{\beta S^1 + \phi_{\text{prev}}} = -1 + \frac{\phi_{\text{prev}} + \gamma}{\beta S^1 + \phi_{\text{prev}}} \tag{2.48}$$

$$I(t) - I(0) = S(0) - S^1(t) + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\beta S^1(t) + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right). \tag{2.49}$$

As $t \to \infty$, we have

$$I_\infty - I(0) = S(0) - S^1_\infty + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\beta S^1_\infty + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right). \tag{2.50}$$

Since $I^1_\infty = 0$ and $S(0) + I(0) = 1 - V(0) = 1 - p$,

$$S^1_\infty = (1-p) + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\beta S^1_\infty + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right) = (1-p) - \frac{\phi_{\text{prev}} + \gamma}{\gamma} R^1_\infty. \tag{2.51}$$

Let

$$w(x) = (1-p) + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\beta x + \phi_{\text{prev}}}{\beta(1-p)(1-\alpha) + \phi_{\text{prev}}}\right). \tag{2.52}$$

We seek solutions to $S^1_\infty = w(S^1_\infty)$ in the range $S^1_\infty \in [0, 1]$. We note that no solutions of equation (2.43) cross the S-nullcline, $S = -\phi_{\text{prev}}/\beta$, and so solutions to equation (2.51) are in the range $[-\phi_{\text{prev}}/\beta, 1]$.

It is possible to use equation (2.51) to find which initial coverage causes solutions of equation (2.43) to hit the $S = 0$ hyper-plane, as they will be those for which $S^1_\infty \leq 0$. As long as this does not occur, equations (2.47), (2.50) and (2.51) are valid also for the modified system (equations (2.44)).

From equation (2.51), we have

$$S^1_\infty = -\frac{1}{\beta}\left(\phi_{\text{prev}} + (\gamma + \phi_{\text{prev}})\, W_i\left(-\frac{\beta S(0) + \phi_{\text{prev}}}{\gamma + \phi_{\text{prev}}} e^{-\frac{\beta(1-p)+\phi_{\text{prev}}}{\gamma+\phi_{\text{prev}}}}\right)\right),$$

where $i = 0$ or -1 specifies the branch of the Lambert W function. To determine which branch of W gives the correct final size, observe that $S^1_\infty < \gamma/\beta$ for any initial condition (in the biologically meaningful domain). This follows from the following argument: If $S(0) \leq \gamma/\beta$, since S is non-increasing and $\dot{S} < 0$ at time $t = 0$ (since also $I(0) > 0$), we are done. If $S(0) > \gamma/\beta$, we note that I is increasing for any S such that $S > \gamma/\beta$. But we have seen that $I_\infty = 0$. Thus, at some point in time, $S < \gamma/\beta$, and since S is non-increasing, we have $S^1_\infty < \gamma/\beta$. From this we see that it is necessary to use the principal branch, W_0 (rather than W_{-1}, which satisfies $W_{-1}(x) \leq -1$ for all x in its domain of

definition, that is $(-1/e, 0)$). Thus,

$$S_\infty^1 = -\frac{1}{\beta}\left(\phi_{\text{prev}} + (\gamma+\phi_{\text{prev}})\, W_0\left(-\frac{\beta S(0)+\phi_{\text{prev}}}{\gamma+\phi_{\text{prev}}} e^{-\frac{\beta(1-p)+\phi_{\text{prev}}}{\gamma+\phi_{\text{prev}}}}\right)\right). \tag{2.53}$$

For convenience, we rewrite equations (2.47), (2.50) and (2.51) to give:

$$\ln\left(\frac{\beta S(0)+\phi_{\text{prev}}}{\beta S_\infty^1+\phi_{\text{prev}}}\right) = \frac{\beta}{\phi_{\text{prev}}+\gamma}\left(1-p-S_\infty^1\right), \tag{2.54}$$

and

$$R_\infty^1 = \frac{\gamma}{\beta}\ln\left(\frac{\beta S(0)+\phi_{\text{prev}}}{\beta S_\infty^1+\phi_{\text{prev}}}\right) = \frac{\gamma}{\phi_{\text{prev}}+\gamma}\left(1-p-S_\infty^1\right). \tag{2.55}$$

Thus,

$$\begin{aligned} V_\infty^1 &= 1 - R_\infty^1 - S_\infty^1 \\ &= 1 - \frac{\gamma}{\phi_{\text{prev}}+\gamma}\left(1-p-S_\infty^1\right) - S_\infty^1 \\ &= \frac{\phi_{\text{prev}}}{\gamma+\phi_{\text{prev}}}(1-S_\infty^1) + \frac{\gamma}{\gamma+\phi_{\text{prev}}}p\,. \end{aligned} \tag{2.56}$$

To see when we can use $S_\infty = S_\infty^1$, $R_\infty = R_\infty^1$ and $V_\infty = V_\infty^1$, it is necessary to find when equation (2.53) yields a negative S_∞^1. First, we evaluate how S_∞ changes with p. To find $\partial_p S_\infty^1$, apply $\partial_p := \frac{\partial}{\partial p}$ to equation (2.51), to get (recall that $S(0) = (1-\alpha)(1-p)$, so $\partial_p S(0) = -(1-\alpha)$)

$$\partial_p S_\infty^1 = \frac{\gamma+\phi_{\text{prev}}}{\beta}\left[\frac{\beta(1-\alpha)}{\beta S(0)+\phi_{\text{prev}}} + \frac{\beta\partial_p S_\infty^1}{\beta S_\infty^1+\phi_{\text{prev}}}\right] - 1 \tag{2.57}$$

$$\left(1 - \frac{\gamma+\phi_{\text{prev}}}{\beta S_\infty^1+\phi_{\text{prev}}}\right)\partial_p S_\infty^1 = \frac{(\gamma+\phi_{\text{prev}})(1-\alpha)}{\beta S(0)+\phi_{\text{prev}}} - 1 \tag{2.58}$$

$$\partial_p S_\infty^1 = \left(\frac{(\gamma+\phi_{\text{prev}})(1-\alpha)}{\beta S(0)+\phi_{\text{prev}}} - 1\right) \Big/ \left(1 - \frac{\gamma+\phi_{\text{prev}}}{\beta S_\infty^1+\phi_{\text{prev}}}\right) \tag{2.59}$$

Since $S_\infty^1 < \gamma/\beta$, it follows that $1 - (\gamma+\phi_{\text{prev}})/(\beta S_\infty^1+\phi_{\text{prev}}) < 0$. Thus,

$$\operatorname{sign}(\partial_p S_\infty^1) = \operatorname{sign}\left[\beta S(0) + \phi_{\text{prev}} - (\gamma+\phi_{\text{prev}})(1-\alpha)\right], \tag{2.60}$$

and since $S(0) = (1-p)(1-\alpha)$, we have:

$$\operatorname{sign}\partial_p S_\infty^1 = \begin{cases} 1 & \text{if } p_m > p \\ 0 & \text{if } p_m = p, \\ -1 & \text{if } p_m < p\,, \end{cases} \tag{2.61}$$

where the local maximum is attained at

$$p = p_m := 1 + \frac{\phi_{\text{prev}} - (\gamma + \phi_{\text{prev}})(1-\alpha)}{\beta(1-\alpha)} = 1 + \frac{\alpha\phi_{\text{prev}} - \gamma(1-\alpha)}{\beta(1-\alpha)}\,. \tag{2.62}$$

Observe that $p_m \in [0,1] \iff \alpha\phi_{\text{prev}} \leq \gamma(1-\alpha)$ and $\alpha\phi_{\text{prev}} \geq (\gamma-\beta)(1-\alpha)$. However, the second condition, which is necessary to ensure $p_m \geq 0$, is trivially satisfied whenever $\beta/\gamma = \mathcal{R}_0 \geq 1$. Thus, if $\mathcal{R}_0 \geq 1$, then $p_m \in [0,1]$ *iff*

$$\alpha\phi_{\text{prev}} \leq \gamma(1-\alpha)\,. \tag{2.63}$$

The maximum value of S^1_∞ is thus:

$$\max_{p\in[0,1]} S^1_\infty = -\frac{1}{\beta}\left(\phi_{\text{prev}} + (\gamma+\phi_{\text{prev}})\, W_0\left(-\frac{\beta S(0) + \phi_{\text{prev}}}{\gamma+\phi_{\text{prev}}} e^{-\frac{\phi_{\text{prev}} - \gamma(1-\alpha)}{(\gamma+\phi_{\text{prev}})(1-\alpha)}}\right)\right)\,. \tag{2.64}$$

Now solving for $S^1_\infty = 0$ using equation (2.53), we have

$$p_0(i) = 1 + \frac{\phi_{\text{prev}}}{\beta(1-\alpha)} + \frac{\gamma+\phi_{\text{prev}}}{\beta}\, W_i\left(-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma+\phi_{\text{prev}})} e^{-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma+\phi_{\text{prev}})}}\right) \tag{2.65}$$

for $i = 0$ or -1. Note that for $\phi_{\text{prev}} > 0$, equation (2.65) gives two values for p_0; we cannot simply cancel out the operation of W with xe^x, since in this case $x = -\frac{\phi_{\text{prev}}}{(1-\alpha)(\phi_{\text{prev}}+\gamma)} < 0$ and W is not univalued for negative arguments. Instead, we have two possibilities for $W\left(\frac{\phi_{\text{prev}}}{(\alpha-1)(\gamma+\phi_{\text{prev}})} e^{\frac{\phi_{\text{prev}}}{(\alpha-1)(\gamma+\phi_{\text{prev}})}}\right)$ corresponding to the two branches, W_0 and W_{-1}.

If $\alpha\phi_{\text{prev}} < (1-\alpha)\gamma$, then $W_0\left(-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma+\phi_{\text{prev}})} e^{-\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma+\phi_{\text{prev}})}}\right) = -\frac{\phi_{\text{prev}}}{(1-\alpha)(\gamma+\phi_{\text{prev}})}$, which gives $p_0(0) = 1$. If $\alpha\phi_{\text{prev}} \geq (1-\alpha)\gamma$ then similarly $p_0(-1) = 1$. This is in agreement with the fact that if $p = 1$, $S(0) = 0$, and so $S^1_\infty = 0$.

There are now three cases, which we express as two main cases, the second of which has two subcases:

- if $p_m \geq 1$ (which happens *iff* $\alpha\phi_{\text{prev}} \geq (1-\alpha)\gamma$), then $S^1_\infty \leq 0 \quad \forall p \in [0,1]$. This follows since if $p_m \geq 1$ then $p_0(-1) = 1 \leq p_m \leq p_0(0)$. Thus because S^1_∞ is increasing for $p < p_m$, so for $p \in [0,1]$, $S^1_\infty \leq S^1_\infty|_{p=1} = 0$. In this case, S_∞ is not given by S^1_∞, but is simply $S_\infty = 0$.
- If $p_m \in (0,1)$ then $S^1_\infty|_{p_m} > S^1_\infty|_{p=1} = 0$ and $p_0(-1) < p_m$.

– If $0 \leq p_0(-1)$ (and $p_0(-1) < p_m < 1$) then

$$S_\infty = \begin{cases} 0 & \text{if } p < p_0(-1), \\ S^1_\infty & \text{if } p_0(-1) \leq p\,. \end{cases} \tag{2.66}$$

– If $p_0(-1) \leq 0$ then $S^1_\infty \geq 0 \quad \forall p \in [0,1]$ and so $S_\infty = S^1_\infty$ for any $p \in [0,1]$.

In all but the very last sub-case, it is also necessary to adjust our formulae for the final sizes of the removed and vaccinated compartments, R_∞ and V_∞, for the values of p for which $S_\infty = 0$. Qualitatively, this adjustment is necessary because, when $\delta \to 0^+$, if a solution reaches $S = 0$ in finite time, S remains 0, while I decays exponentially to 0. However, the solutions of equations (2.43) are only identical to those of equations (2.44) so long as $S > 0$. Moreover, once $S = 0$, V remains constant and all the infectives move into the recovered compartment, which is not the case for solutions of equations (2.43).

To find formulas for R_∞ and V_∞, fix $p \in [0,1)$ and let t_0 be the first time at which no susceptibles remain ($S(t_0) = S^1(t_0) = 0$). Then equations (2.43) are valid for any $t < t_0$. We now have

$$\begin{aligned} R(t) &= \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\beta S(t) + \phi_{\text{prev}}}\right) \\ I(t) &= I(0) + S(0) - S(t) + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\beta S(t) + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right), \end{aligned}$$

in a manner analogous to equation (2.47) and equation (2.50). These equations depend on $S = S(t)$ in a way that is continuous at $S = 0$, and so taking $t \to t_0$ is equivalent to taking $S \to 0$:

$$R(t_0) = \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\phi_{\text{prev}}}\right) \tag{2.67}$$

$$I(t_0) = I(0) + S(0) + \frac{\phi_{\text{prev}} + \gamma}{\beta} \ln\left(\frac{\phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right). \tag{2.68}$$

When $S = 0$, I decays exponentially to 0, until all the infectives present at t_0 transition into the removed class, R. Thus, we have

$$R_\infty = R(t_0) + I(t_0) = 1 - p - \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right). \tag{2.69}$$

Next, we know that once $S = 0$, V does not change either, since there are no more

susceptibles to be vaccinated. Thus,

$$V_\infty = V(t_0) = 1 - R(t_0) - I(t_0) = 1 - R_\infty = p + \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right) \tag{2.70}$$

To summarize our results so far,

$$S_\infty = \begin{cases} 0 & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ S_\infty^1 & \text{if } p_0 \leq p \leq 1\,, \end{cases} \tag{2.71a}$$

$$R_\infty = \begin{cases} 1 - p - \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right) & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ \frac{\gamma}{\gamma + \phi_{\text{prev}}} \left(1 - p - S_\infty^1\right) & \text{if } p_0 \leq p \leq 1\,, \end{cases} \tag{2.71b}$$

$$V_\infty = \begin{cases} p + \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right) & \text{if } p < p_0 \text{ or } 1 \leq p_m, \\ \frac{1}{\gamma + \phi_{\text{prev}}} \left(\phi_{\text{prev}}(1 - S_\infty^1) + \gamma p\right) & \text{if } p_0 \leq p \leq 1\,, \end{cases} \tag{2.71c}$$

where S_∞^1 is given by equation (2.53).

Qualitative behaviour of S_∞^1 for high vaccine coverage

Qualitatively, observe that for high values of p, $S_\infty^1 \approx (1-\alpha)(1-p)$ (see figure 2.E.1). This is because when $p > p_d := 1 - \frac{\gamma}{\beta(1-\alpha)}$ then $S(0) < \gamma/\beta$, in which case I decays to 0 monotonically (the subscript "d" denotes decay of I). Now in this case, S decreases at least as fast as it does in the vaccination-less SIR model, and so I decreases at least as fast as in the vaccination-less case too. Because S is monotonically decreasing, I decays faster than $I(0)e^{(\beta S(0) - \gamma)t}$. Because I decays at least exponentially, S hardly changes over the course of the epidemic, and so we see an approximately linear decay in $S_\infty^1 \approx S(0) = (1-\alpha)(1-p)$ to 0 as we increase p. This last phenomenon is related to the herd immunity effect in the standard SIR model: when the entire population is susceptible, in the absence of post-outbreak vaccination ($\phi_{\text{prev}} = 0$), the critical vaccine coverage which stops the epidemic from taking off is $p = 1 - \frac{1}{\mathcal{R}_{\text{eff}}} = 1 - \frac{\gamma}{\beta(1-\alpha)}$. But note, moreover, that the maximal value of S_∞^1 is not attained at p_d. Rather, we see that

$$p_m - p_d = \frac{\alpha(\phi_{\text{prev}} + \gamma)}{\beta(1-\alpha)} > 0\,. \tag{2.72}$$

This is because even when $p > p_d$ and I immediately decays to 0 (starting at $t = 0$), there are still some susceptibles converted into vaccinated individuals, due to $I > 0$. This number of susceptibles lost to vaccination decreases as p is increased, since this decreases $I(0)$ as well. Thus, for $p > p_d$, initially S_∞ increases with p. Only when $p > p_m$ does the decrease in S_∞ due to more susceptibles being vaccinated pre-emptively take over.

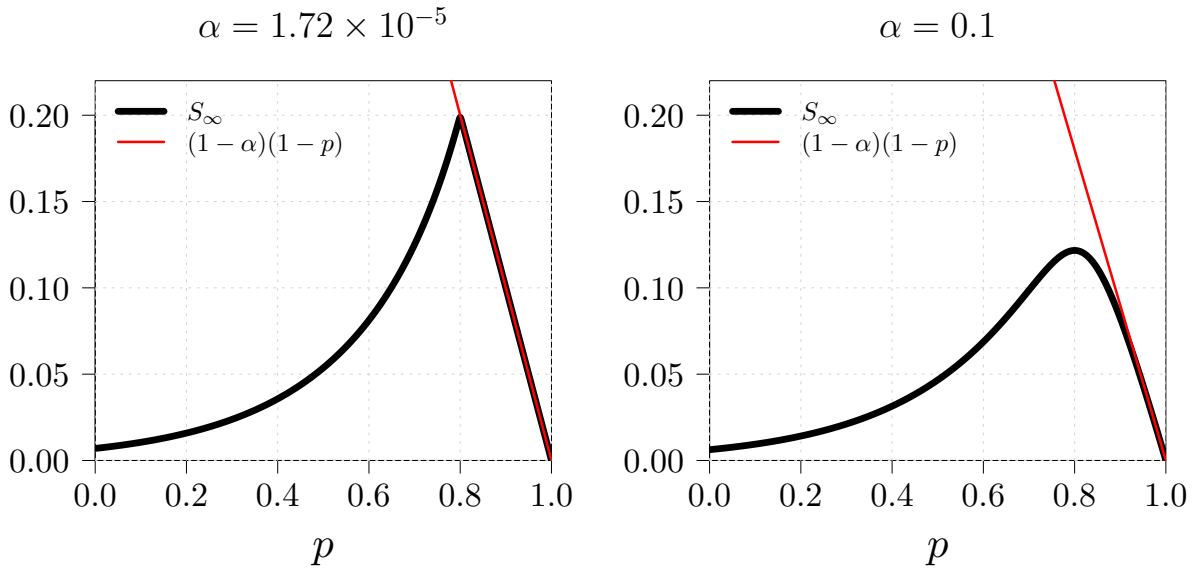

Figure 2.E.1: The proportion of individuals remaining susceptible at the end of the epidemic, S_∞, as a function of the proportion of the population vaccinated pre-emptively, p, for the model in which vaccination rate is proportional to prevalence. The line $y(p) = (1-\alpha)(1-p)$ is overlaid in red. We take the proportion of susceptibles initially infected, α, to be the estimated value in the left panel ($\alpha = 1.72 \times 10^{-5}$) and a much larger value for comparison in the right panel ($\alpha = 0.1$); the remaining model parameters are as in table 2.1. See §2.E.1.

π_p decreases and ψ_p increases with p when $p_m \geq 1$

Consider the first case above: if $p_m \geq 1$ (see equation (2.71)), then $S_\infty = 0$ and we have

$$V_\infty = p + \frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right)$$
$$\psi_p = \frac{\phi_{\text{prev}}}{\beta(1-p)} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right).$$

Note that since $S_\infty = 0$, $\pi_p = 1 - \psi_p$. Letting $x = \frac{\beta}{\phi_{\text{prev}}} S(0) = \frac{\beta(1-\alpha)}{\phi_{\text{prev}}}(1-p)$,

$$\psi(x) = \psi_{p(x)} = (1-\alpha)\frac{\ln(x+1)}{x},$$

so

$$\begin{aligned}
\partial_p \psi_p &= -\frac{\beta(1-\alpha)}{\phi_{\text{prev}}} \frac{\partial \psi(x)}{\partial x} = -\frac{\beta(1-\alpha)}{\phi_{\text{prev}}} \frac{x - (1+x)\ln(1+x)}{x^2(x+1)} \\
&= -\frac{\beta(1-\alpha)}{\phi_{\text{prev}}} \frac{\frac{\beta}{\phi_{\text{prev}}} S(0) - (\frac{\beta}{\phi_{\text{prev}}} S(0) + 1)\ln(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1)}{\left(\frac{\beta}{\phi_{\text{prev}}} S(0)\right)^2 (\frac{\beta}{\phi_{\text{prev}}} S(0) + 1)} \\
&= -\frac{\phi_{\text{prev}}(1-\alpha)}{\beta} \frac{\beta S(0) - (\beta S(0) + \phi_{\text{prev}})\ln(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1)}{(S(0))^2 (\beta S(0) + \phi_{\text{prev}})}.
\end{aligned} \tag{2.73}$$

Since $x < (x+1)\ln(1+x) \quad \forall x > -1$, it follows that for $p \in [0,1)$, $\partial_p \psi_p > 0$ and consequently, $\partial_p \pi_p < 0$.

Behaviour of $p_0(-1)$ as $\phi_{\text{prev}} \to 0^+$

Using equation (2.65), one can show that $\lim_{\phi_{\text{prev}} \to 0^+} p_0(-1) = -\infty$. This follows because $\lim_{\phi_{\text{prev}} \to 0^-} W_{-1}(x) = -\infty$. But because $p_0(-1) \to 1$ as $\phi_{\text{prev}} \to \frac{\gamma(1-\alpha}{\alpha}$, it follows that there is some value of ϕ_{prev} for which $p_0(-1) = 0$. This ϕ_{prev} can be found from equation (2.65), but the formula is not needed here.

However, we also note that $\phi_{\text{prev}} < \frac{\gamma(1-\alpha}{\alpha}$ (along with $\mathcal{R}_0 > 1$) implies $p_m \in (0,1)$, so for small ϕ_{prev}, we have $p_m \in (0,1)$ and $p_0(-1) < 0$ (in fact, as $\phi_{\text{prev}} \to 0^+$, $p_m \to 1 - 1/\mathcal{R}_0$). Thus, for $\phi_{\text{prev}} \to 0^+$ (small enough such that $p_0(-1) < 0$), $S_\infty^1 > 0$ for $p \in [0,1)$.

2.E.2 Vaccination rate $\propto$ incidence

The model equations are

$$\dot{S} = -\beta SI - \phi_{\text{inc}} SI \tag{2.74a}$$
$$\dot{I} = \beta SI - \gamma I \tag{2.74b}$$
$$\dot{R} = \gamma I \tag{2.74c}$$
$$\dot{V} = \phi_{\text{inc}} SI\,. \tag{2.74d}$$

Finding the final sizes for this model is somewhat similar to when vaccination is proportional to prevalence:

$$\frac{\mathrm{d}I}{\mathrm{d}S} = -\frac{\beta SI - \gamma I}{(\beta + \phi_{\text{inc}})SI} = \frac{\gamma}{\beta + \phi_{\text{inc}}}\frac{1}{S} - \frac{\beta}{\beta + \phi_{\text{inc}}} \tag{2.75}$$
$$I(t) - I(0) = \frac{\gamma}{\beta + \phi_{\text{inc}}} \ln\left(\frac{S(t)}{S(0)}\right) - \frac{\beta}{\beta + \phi_{\text{inc}}}\left(S(t) - S(0)\right) \tag{2.76}$$
$$-(1-p)\alpha = \frac{\gamma}{\beta + \phi_{\text{inc}}} \ln\left(\frac{S_\infty}{(1-p)(1-\alpha)}\right) - \frac{\beta}{\beta + \phi_{\text{inc}}}\left(S_\infty - (1-p)(1-\alpha)\right) \tag{2.77}$$
$$S_\infty = \frac{\gamma}{\beta} \ln\left(\frac{S_\infty}{(1-p)(1-\alpha)}\right) + \left(1 + \frac{\phi_{\text{inc}}}{\beta}\alpha\right)(1-p)\,, \tag{2.78}$$

where equation (2.77) is obtained by taking $t \to \infty$ in equation (2.76). The solution of equation (2.78) is given explicitly by

$$S_\infty = -\frac{\gamma}{\beta}\, W_0\left(-\frac{\beta(1-p)(1-\alpha)}{\gamma} e^{-\frac{\beta + \phi_{\text{inc}}\alpha}{\gamma}(1-p)}\right)\,, \tag{2.79}$$

where we take the principle branch of the Lambert function, W_0, because solutions are in the range $S_\infty \in [0, \gamma/\beta]$ (see equation (2.33) in appendix 2.C). Note also that $S_\infty > 0$ *iff* $S(0) > 0$ (that is, $p < 1$ and $\alpha < 1$). Thus, $S = 0$ is not attainable in scenarios of interest here.

To find V_∞, we proceed similarly:

$$\dot{V} = -\frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}\dot{S} \tag{2.80}$$
$$V_\infty - V(0) = -\frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}(S_\infty - S(0)) \tag{2.81}$$
$$V_\infty - p = \frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}\big((1-p)(1-\alpha) - S_\infty\big)\,. \tag{2.82}$$

At the end of the epidemic, $1 = R_\infty + S_\infty + V_\infty$, hence,

$$\begin{aligned} R_\infty &= 1 - p + \frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}(S_\infty - (1-p)(1-\alpha)) - S_\infty \\ &= 1 - p - \frac{\phi_{\text{inc}}(1-p)(1-\alpha) + \beta S_\infty}{\beta + \phi_{\text{inc}}} \,. \end{aligned} \tag{2.83}$$

Thus,

$$\pi_p = \frac{R_\infty}{1-p} = 1 - \frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}(1-\alpha) - \frac{\beta S_\infty}{(\beta + \phi_{\text{inc}})(1-p)} \tag{2.84a}$$

$$\psi_p = \frac{V_\infty - p}{1-p} = \frac{\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}}\left((1-\alpha) - \frac{S_\infty}{1-p}\right) . \tag{2.84b}$$

Note that whenever $S(0) > 0$, there are susceptible individuals left at the end of the epidemic, and so $\pi_p \neq 1 - \psi_p$.

Using $\partial_p \left(\frac{S_\infty}{1-p}\right) = \frac{(1-p)\partial_p S_\infty + S_\infty}{(1-p)^2}$ we have

$$\begin{aligned} \partial_p S_\infty &= \frac{\gamma}{\beta}\left(\frac{\partial_p S_\infty}{S_\infty} + \frac{1}{1-p}\right) - \left(1 + \frac{\phi_{\text{inc}}}{\beta}\alpha\right) \\ \frac{\beta S_\infty - \gamma}{\beta S_\infty}\partial_p S_\infty &= \frac{\gamma}{\beta(1-p)} - 1 - \frac{\phi_{\text{inc}}}{\beta}\alpha \\ \partial_p S_\infty &= \frac{\gamma}{1-p}\frac{S_\infty}{\beta S_\infty - \gamma} - (\phi_{\text{inc}}\alpha + \beta)\frac{S_\infty}{\beta S_\infty - \gamma} \\ &= \frac{\gamma - (\phi_{\text{inc}}\alpha + \beta)(1-p)}{(1-p)}\frac{S_\infty}{\beta S_\infty - \gamma} \\ (1-p)\partial_p S_\infty + S_\infty &= \frac{(\beta S_\infty - (1-p)(\beta + \phi_{\text{inc}}\alpha))S_\infty}{\beta S_\infty - \gamma} \\ \partial_p\left(\frac{S_\infty}{1-p}\right) &= \frac{(\beta S_\infty - (1-p)(\beta + \phi_{\text{inc}}\alpha))S_\infty}{\beta S_\infty - \gamma}\frac{1}{(1-p)^2} \,. \end{aligned} \tag{2.85}$$

From equation (2.78)

$$\beta S_\infty - (1-p)(\beta + \phi_{\text{inc}}\alpha) = \gamma \ln\left(\frac{S_\infty}{S(0)}\right) < 0 \tag{2.86}$$

because $S_\infty < S(0)$ (unless $S(0)$ or $I(0)$ is 0, in which case an outbreak cannot take place).

Thus $\partial_p \left(\frac{S_\infty}{1-p} \right) > 0$ and so

$$\partial_p \pi_p < 0 \tag{2.87}$$
$$\partial_p \psi_p < 0 \,. \tag{2.88}$$

Note also that S_∞ attains a local maximum (in p) at

$$p_m = 1 - \gamma/(\phi_{\text{inc}}\alpha + \beta) \,. \tag{2.89}$$

$p_m < 0$ for $\gamma > \phi_{\text{inc}}\alpha + \beta$, in which case S_∞ decreases with p on the interval $[0, 1]$. This can only happen when $\gamma > \beta$, that is when $\mathcal{R}_0 < 1$, implying that for any disease which can spread in the population (with no vaccination), pre-emptive vaccination initially raises, then lowers the proportion of susceptibles remaining at the end of the epidemic. The maximum level of remaining susceptibles is

$$S_\infty|_{p=p_m} = -\frac{\gamma}{\beta} \, W_0 \left(-\frac{\beta(1-\alpha)}{(\phi_{\text{inc}}\alpha + \beta)e} \right) \,. \tag{2.90}$$

However, R_∞ is more informative, since susceptibles can be depleted by either infection or vaccination, and so fewer remaining susceptibles does not necessarily imply a larger epidemic, nor does it imply that more individuals were vaccinated. However,

$$\partial_p R_\infty = \partial_p \left((1-p)\pi_p \right) = -\pi_p + (1-p)\partial_p \pi_p < 0 \,, \tag{2.91}$$

which shows that increasing pre-emptive vaccine coverage decreases the size of the epidemic, as expected.

2.E.3 Vaccination rate $\propto$ proportion still susceptible

The model equations are

$$\dot{S} = -\beta SI - \phi_{\text{susc}} S \tag{2.92a}$$
$$\dot{I} = \beta SI - \gamma I \tag{2.92b}$$
$$\dot{R} = \gamma I \tag{2.92c}$$
$$\dot{V} = \phi_{\text{susc}} S \,. \tag{2.92d}$$

In this case, a similar strategy to the one we employed for the case where vaccination is proportional to prevalence doesn't quite work. Calculating $S_\infty(I(0))$ is not enough, since we do not know how the remainder of the population is partitioned between the removed and vaccinated classes at the end of the epidemic. The following calculations are also

helpful but insufficient:

$$\frac{\dot{I}}{I} = \beta S - \gamma = \frac{\beta}{\phi_{\text{susc}}}\dot{V} - \gamma \tag{2.93}$$

$$\frac{\dot{S}}{S} = -\beta I - \phi_{\text{susc}} = -\frac{\beta}{\gamma}\dot{R} - \phi_{\text{susc}}\,, \tag{2.94}$$

so

$$\ln\left(\frac{I(t)}{I(0)}\right) = \frac{\beta}{\phi_{\text{susc}}}(V(t) - V(0)) - \gamma t \tag{2.95}$$

$$\ln\left(\frac{S(t)}{S(0)}\right) = -\frac{\beta}{\gamma}(R(t) - R(0)) - \phi_{\text{susc}} t\,. \tag{2.96}$$

However, it is not possible to extract V_∞ and R_∞ from here because phase-portrait arguments show that I and S tend to 0 as $t \to \infty$ (this is also implied by equations (2.95) and (2.96)), thus both sides of these equations diverge as $t \to \infty$.

Nonetheless, a similar method to the one employed in appendix 2.E.2 yields a relation between $S(t)$ and $I(t)$ from which, using the previous relations, a relation between $R(t)$ and $V(t)$ can be obtained. These will not diverge as $t \to \infty$ (they are bounded), so any divergent components must cancel out.

$$\frac{\mathrm{d}S}{\mathrm{d}I} = -\frac{(\beta I + \phi_{\text{susc}})S}{(\beta S - \gamma)I} \tag{2.97}$$

$$\frac{(\beta S - \gamma)}{S}\mathrm{d}S = -\frac{(\beta I + \phi_{\text{susc}})}{I}\mathrm{d}I \tag{2.98}$$

$$(\beta - \frac{\gamma}{S})\mathrm{d}S = -(\beta + \frac{\phi_{\text{susc}}}{I})\mathrm{d}I \tag{2.99}$$

$$\beta(S(t) - S(0)) - \gamma \ln\left(\frac{S(t)}{S(0)}\right) = -\beta(I(t) - I(0)) - \phi_{\text{susc}} \ln\left(\frac{I(t)}{I(0)}\right) \tag{2.100}$$

from which we get:

$$S(t) = -\frac{\gamma}{\beta}\, W\left(-\frac{\beta}{\gamma} e^{\frac{\beta}{\gamma}(I(t)-I(0)-S(0))} (\frac{I(t)}{I(0)})^{\frac{\phi_{\text{susc}}}{\gamma}} S(0)\right)\,. \tag{2.101}$$

By taking the limit $t \to \infty$ in equation (2.101), we see that $S_\infty = 0$, and conseqently, $\psi_p = 1 - \pi_p$.

We now determine under which conditions each of the two branches of the Lambert W

function is used in equation (2.101). First, note that

$$\begin{aligned} S(I) &= -\frac{\gamma}{\beta}\, W(z(p)) \\ z(I) &= -\frac{\beta}{\gamma} e^{\frac{\beta}{\gamma}(I(t)-I(0)-S(0))} \left(\frac{I(t)}{I(0)}\right)^{\frac{\phi_{\text{susc}}}{\gamma}} S(0) \\ z(I(0)) &= -\frac{\beta}{\gamma} S(0) e^{-\frac{\beta}{\gamma}S(0)}\,. \end{aligned}$$

For $I = I(0)$, we expect to get $W(-\frac{\beta}{\gamma}S(0)) = -\frac{\beta}{\gamma}S(0)$ so that $S(I = I(0)) = S(t = 0)$. We know that $S(0) > \frac{\gamma}{\beta}$ which implies that for $t = 0$, we must use W_1. Because $S(t)$ monotonically decreases to 0 as $t \to 0$, we know that the branch W_1 is used until the peak prevalence is attained (at which time $S = \gamma/\beta$), and then the principal branch W_0 is used. Now,

$$\begin{aligned} \frac{\mathrm{d}R}{\mathrm{d}I} &= \frac{\gamma I}{\beta S I - \gamma I} \\ &= \frac{\gamma}{\beta S - \gamma} \\ &= \frac{-1}{W\left(-\frac{\beta}{\gamma} e^{\frac{\beta}{\gamma}(I-I(0)-S(0))} (\frac{I}{I(0)})^{\frac{\phi_{\text{susc}}}{\gamma}} S(0)\right) + 1}\,. \end{aligned} \tag{2.102}$$

This can be integrated, to give

$$R_\infty = R(0) + \int_{I(0)}^{I_\infty} \frac{\gamma}{\beta S - \gamma}\, \mathrm{d}I \tag{2.103}$$

$$= R(0) + \int_{I(0)}^{I_\infty} \frac{-1}{W_i\left(-\frac{\beta}{\gamma} e^{\frac{\beta}{\gamma}(I-I(0)-S(0))} (\frac{I}{I(0)})^{\frac{\phi_{\text{susc}}}{\gamma}} S(0)\right) + 1}\, \mathrm{d}I$$

$$= \int_0^{I(0)} \frac{1}{W_i\left(-\frac{\beta}{\gamma} e^{\frac{\beta}{\gamma}(I-I(0)-S(0))} (\frac{I}{I(0)})^{\frac{\phi_{\text{susc}}}{\gamma}} S(0)\right) + 1}\, \mathrm{d}I\,, \tag{2.104}$$

where the appropriate branch of W_i is determined as above, as the integration varible I is varied. Note, however, that the integral in equation (2.103) is improper, because the integral diverges at the peak prevalence (when $S = \gamma/\beta$).

2.E.4 Instantaneous vaccination of a proportion ϕ_{inst} of the population

In this case, the disease progresses according to the standard SIR model,

$$\dot{S} = -\beta SI\,, \tag{2.105a}$$
$$\dot{I} = (\beta S - \gamma)I\,, \tag{2.105b}$$
$$\dot{R} = \gamma I\,, \tag{2.105c}$$

with initial conditions given by

$$\begin{aligned} S(0) &= (1-p)(1-\alpha)(1-\phi_{\text{inst}}) \\ I(0) &= (1-p)\alpha \\ R(0) &= 0 \\ V(0) &= p + \phi_{\text{inst}}(1-p)(1-\alpha)\,. \end{aligned}$$

Note that for this model, $S(0)$ is the density of susceptibles after the post-outbreak vaccination response has taken place.

Equation (2.105a) implies that

$$-\frac{\gamma}{\beta}\frac{\mathrm{d}}{\mathrm{d}t}\ln(S) = \dot{R}\,, \tag{2.106}$$

thus S_∞ satisfies the equation

$$\frac{\gamma}{\beta}\ln\left(\frac{S(0)}{S_\infty}\right) = R_\infty = 1 - V(0) - S_\infty\,, \tag{2.107}$$

or

$$S_\infty = -\frac{\gamma}{\beta}\,W_0\left(-\frac{\beta}{\gamma}S(0)e^{-\frac{\beta}{\gamma}(1-V(0))}\right)\,. \tag{2.108}$$

We use the principle branch of the Lambert function in order to obtain solutions satisfying $S_\infty \leq S(0)$. Since $-\frac{\beta}{\gamma}S(0)e^{-\frac{\beta}{\gamma}(1-V(0))} > -\frac{\beta}{\gamma}S(0)e^{-\frac{\beta}{\gamma}S(0)}$, and W_1 is monotonically decreasing, $-\frac{\gamma}{\beta}\,W_1\left(-\frac{\beta}{\gamma}S(0)e^{-\frac{\beta}{\gamma}(1-V(0))}\right) > S(0)$, which does not correspond to biologically feasible solutions. In addition, we have $R_\infty = \frac{\gamma}{\beta}\ln\frac{S(0)}{S_\infty}$.

Since there is no vaccination except during the initial (immediate) response to the out-

break, $V_\infty = V(0) = p + \phi_{\text{inst}}(1-p)(1-\alpha)$, and so

$$\psi_p = \frac{V_\infty - p}{1-p} = \phi_{\text{inst}}(1-\alpha) \tag{2.109a}$$

$$\pi_p = 1 - \psi_p - \frac{S_\infty}{1-p}\,. \tag{2.109b}$$

Using equation (2.107), we also have

$$\pi_p = -\frac{\gamma}{\beta(1-p)} \ln\left(\frac{1-\pi_p-\phi_{\text{inst}}(1-\alpha)}{(1-\alpha)(1-\phi_{\text{inst}})}\right) \tag{2.110}$$

It follows from equation (2.109b) that π_p is a decreasing function of p: from equation (2.107) we have

$$\frac{\gamma}{\beta}\left(-\frac{1}{1-p} - \frac{\partial_p S_\infty}{S_\infty}\right) = -1 + \phi_{\text{inst}}(1-\alpha) - \partial_p S_\infty \tag{2.111}$$

and so

$$\partial_p S_\infty = \left(-1 + \phi_{\text{inst}}(1-\alpha) + \frac{\gamma}{\beta}\frac{1}{1-p}\right)\left(\frac{\beta S_\infty}{\beta S_\infty - \gamma}\right)\,. \tag{2.112}$$

This gives

$$\begin{aligned}(1-p)\partial_p S_\infty + S_\infty &= \frac{\beta S_\infty}{\beta S_\infty - \gamma}\left((1-p)\left(-1+\phi_{\text{inst}}(1-\alpha)\right) + S_\infty\right) \\ &= \frac{\beta S_\infty}{\gamma - \beta S_\infty}\left(1 - V_\infty - S_\infty\right)\,,\end{aligned} \tag{2.113}$$

which is positive so long as $R_\infty > 0$ (this happens when $S(0) > S_\infty$, which is true whenever $S(0)$ and $I(0)$ are not 0). It now follows that $\partial_p \pi_p = -\frac{(1-p)\partial_p S_\infty + S_\infty}{(1-p)^2} < 0$. Note that the probability of a delayer being vaccinated post-outbreak (ψ_p) is constant.

Lastly, note that

$$(1-p)\partial_p S_\infty = \left(1 - V_\infty - \frac{\gamma}{\beta}\right)\left(\frac{\beta S_\infty}{\gamma - \beta S_\infty}\right)\,, \tag{2.114}$$

which implies that S_∞ increases with p *iff* $(1-p)(1-\phi_{\text{inst}}(1-\alpha))\beta > \gamma$, or equivalently, $I(0) + S(0) = 1 - V_\infty > \frac{\gamma}{\beta}$. Compare this to the more stringent condition $S(0) > \frac{\gamma}{\beta}$ which ensures that the epidemic takes off ($I'(0) > 0$).

2.F Maximal vaccination rate for fair comparison of models

In this section, we find the fair comparison values of the vaccination efforts ϕ_{inc} and ϕ_{prev} (see §2.7.2). These are defined as the levels of vaccination effort ϕ_{inc} and ϕ_{prev} that result in maximal vaccination rates equal to 0.1/day (that is, comparable to [37]).

2.F.1 Maximal Vaccination rate when $\dot{V} = \phi_{\text{prev}} I$

We begin by finding what the maximal vaccination is when $\dot{V} = \phi_{\text{prev}} I$. Because the vaccination rate, $\dot{V}$, is maximal when prevalence, I, is maximal, we aim to find the maximal prevalence. Now observe that since $\dot{I} = (\beta S - \gamma) I$, incidence is maximal when $S = \gamma/\beta$. Thus, the peak prevalence is found by substituting $S = \gamma/\beta$, into equation (2.49) to obtain

$$I_{\text{peak}} = 1 - p - \gamma/\beta + \frac{\gamma + \phi_{\text{prev}}}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right) \tag{2.115}$$

(recall that $I(0) + S(0) = 1 - p$). We now wish to find at which value of p the maximal vaccination rate (over time) is largest. Observe that

$$\frac{\partial}{\partial p} I_{\text{peak}} = -1 + \frac{(\gamma + \phi_{\text{prev}})(1 - \alpha)}{\beta S(0) + \phi_{\text{prev}}}, \tag{2.116}$$

$$\frac{\partial^2}{\partial p^2} I_{\text{peak}} = \frac{\beta(\gamma + \phi_{\text{prev}})(1 - \alpha)^2}{(\beta S(0) + \phi_{\text{prev}})^2} > 0, \tag{2.117}$$

which implies that the peak prevalence (and thus the maximal vaccination rate) has a minium in p when

$$p_{\text{crit}} = \frac{(\beta - \gamma)(1 - \alpha) + \phi_{\text{prev}}\alpha}{\beta(1 - \alpha)} > 0 \tag{2.118}$$

($\beta > \gamma$ because $\mathcal{R}_0 > 1$).

There are now two possibilities:

- If $p_{\text{crit}} \geq 1$ (which happens *iff* $\phi_{\text{prev}}\alpha \leq \gamma(1 - \alpha)$) then the maximal vaccination rate is attained when $p = 0$.
- If $p_{\text{crit}} < 1$ (which happens *iff* $\phi_{\text{prev}}\alpha > \gamma(1 - \alpha)$) then, the maximal vaccination rate must be attained either when $p = 0$ or when $p = 1$.

Noting that

$$I_{\text{peak}}|_{p=0} = 1 - \gamma/\beta + \frac{\gamma + \phi_{\text{prev}}}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\beta(1-\alpha) + \phi_{\text{prev}}}\right)$$
$$I_{\text{peak}}|_{p=1} = -\gamma/\beta + \frac{\gamma + \phi_{\text{prev}}}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\phi_{\text{prev}}}\right), \tag{2.119}$$

it follows that

$$\dot{V}_{\max} = -\phi_{\text{prev}}\frac{\gamma}{\beta} + \phi_{\text{prev}} \max\left\{1 + \frac{\gamma + \phi_{\text{prev}}}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\beta(1-\alpha) + \phi_{\text{prev}}}\right), \right.$$
$$\left. \frac{\gamma + \phi_{\text{prev}}}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\phi_{\text{prev}}}\right)\right\}. \tag{2.120}$$

We also note that

$$\frac{\partial}{\partial \phi_{\text{prev}}} I_{\text{peak}} = \frac{1}{\beta} \ln\left(\frac{\gamma + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right) + \frac{1}{\beta}\left(1 - \frac{\gamma + \phi_{\text{prev}}}{\beta S(0) + \phi_{\text{prev}}}\right). \tag{2.121}$$

Because $1 - x + \ln(x) < 0$ for any $0 < x \neq 1$, it follows that the peak prevalence I_{peak}, and thus the peak vaccination rate, decreases with increasing vaccination effort, ϕ_{prev} (for any initial coverage, p). Also, as $\phi_{\text{prev}} \to \infty$, we have

$$I_{\text{peak}}|_{p=0} \to \alpha\,,$$
$$I_{\text{peak}}|_{p=1} \to 0\,.$$

Setting $\dot{V}_{\max} = 0.1/\text{day}$ in equation (2.120) we can numerically solve for ϕ_{prev}, with α, β, γ, as in Tables 2.1 and 2.2 to obtain $\phi_{\text{prev}} \approx 1582/\text{day}$.

2.F.2 Maximal Vaccination rate when $\dot{V} = \phi_{\text{inc}} SI$

First, we will derrive a formula for the maximal vaccination rate as it depends on the model parameters, α, β, γ, ϕ_{inc}. We then use this formula to calculate the appropriate range for ϕ_{inc}, given the estimates of the other parameters cited in Tables 2.1 and 2.2.

Differentiating equation (2.74d), we have

$$\ddot{V} = \phi_{\text{inc}}\left(\dot{S}I + S\dot{I}\right) = \phi_{\text{inc}} SI\left(\beta S - \gamma - (\beta + \phi_{\text{inc}})I\right).$$

Thus, critical points of $\dot{V}$ (excluding those for which $\dot{V} = 0$) occur when $\beta S - \gamma =$

$(\beta + \phi_{\text{inc}})I$. Using equation (2.76) and simplifying, this is equivalent to

$$2\beta S = \gamma \ln S + \phi_{\text{inc}}(1-p)\alpha + \beta(1-p) + \gamma\Big(1 - \ln\big((1-p)(1-\alpha)\big)\Big)\,,$$

which has two formal solutions,

$$\hat{S}_k = -\frac{\gamma}{2\beta}\, W_k\left(-2\frac{\beta}{\gamma}(1-p)(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}(1-p)-1}\right)\,,$$

with $k = 0$ or -1.

However, it is impossible for $\hat{S}_0$ to be attained by $S(t)$, for all $t \geq 0$. To see this, suppose, in order to derive a contradiction, that there is some time $\hat{t}_0 \geq 0$ such that $S(\hat{t}_0) = \hat{S}_0$. Note that $-1 \leq W_0 < 0$ on the interval $[-1/e, 0)$, so $0 < \hat{S}_0 < \gamma/\beta$. Because $\hat{S}_0 < \gamma/\beta$, we have $\dot{I}(\hat{t}_0) < 0$. Since $\dot{S} < 0$ (for all time t), it follows that $\ddot{V} < 0$ when $t = \hat{t}_0$, in contradiction to the fact that by definition of $\hat{S}_0$, $\ddot{V}(\hat{t}_0) = 0$. Thus, $S(t) > \hat{S}_0$ is proven[1]. It follows that if there is a biologically relevant value of S at which $\ddot{V}$ changes signs, it must be

$$\hat{S}_{-1} = -\frac{\gamma}{2\beta}\, W_{-1}\left(-2\frac{\beta}{\gamma e}(1-p)(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}(1-p)}\right)\,. \tag{2.122}$$

Note that $\hat{S}_{-1} > \frac{\gamma}{2\beta}$ because $W_{-1}(x) < -1 \quad \forall x \in [-1/e, 0]$ (but it is also possible that $\hat{S}_{-1} > S(0) = (1-p)(1-\alpha)$, which would make this critical point biologically unfeasible).

Because S decreases with time, we see that $\ddot{V}$ can change signs at most once for all $t \geq 0$. Observe that since $0 < S_\infty < \gamma/\beta$, and $\dot{V}$ decreases when $S \in (0, \gamma/\beta]$, it follows that $\dot{V}$ eventually (*i.e.,* for large enough t) decreases with time. Hence, if $\ddot{V}(0) \leq 0$, then $t = 0$ is a maximum of $\dot{V}$ for $t \geq 0$, and if $\ddot{V}(0) > 0$ then $\dot{V}$ attains its maximum when $S(t) = \hat{S}_{-1}$. The sign of $\ddot{V}(0)$ is identical to the sign of $\beta S(0) - \gamma - (\beta + \phi_{\text{inc}})I(0)$, so the maximal vaccination rate is attained at $t = 0$ if $(1-p)\big((1-2\alpha)\beta - \alpha\phi_{\text{inc}}\big) \leq \gamma$, and when $S(t) = \hat{S}_{-1}$ otherwise. Thus,

$$\nu(p) = \max_{t\geq 0} \dot{V} = \begin{cases} \phi_{\text{inc}}(1-p)^2(1-\alpha)\alpha & \text{if } (1-p)((1-2\alpha)\beta - \alpha\phi_{\text{inc}}) \leq \gamma\,, \\ \frac{\phi_{\text{inc}}}{\phi_{\text{inc}}+\beta}(\beta\hat{S}_{-1} - \gamma)\hat{S}_{-1} & \text{if } (1-p)((1-2\alpha)\beta - \alpha\phi_{\text{inc}}) > \gamma\,, \end{cases} \tag{2.123}$$

(where, for the second case, we used the fact that $I = \frac{\beta\hat{S}_{-1}-\gamma}{\phi_{\text{inc}}+\beta}$ when $S = \hat{S}_{-1}$).

To maximize ν over all $p \in [0, 1)$ (with α, β, γ and ϕ_{inc} fixed), we consider the following 3 cases:

[1]Consequently, $S_\infty \geq \hat{S}_0$, which is equivalent to a statement about the Lambert W function: $W_0(2x/e) \geq 2\, W_0(x)$, for $-1/e < x < 0$.

- First, if $0 < \gamma < (1-2\alpha)\beta - \alpha\phi_{\text{inc}}$, then $(1-p)((1-2\alpha)\beta - \alpha\phi_{\text{inc}}) \leq \gamma$ is equivalent to $\hat{p} = 1 - \frac{\gamma}{(1-2\alpha)\beta - \alpha\phi_{\text{inc}}} \leq p$, and $\hat{p} \in (0,1)$. Hence, for $p \in [\hat{p}, 1)$, $\nu(p) = \phi_{\text{inc}}(1-p)^2(1-\alpha)\alpha$ is a decreasing function of p, and so $\max_{p\in[\hat{p},1)} \nu(p) = \phi_{\text{inc}}\left(\frac{\gamma}{(1-2\alpha)\beta - \alpha\phi_{\text{inc}}}\right)^2(1-\alpha)\alpha$, and is attained when $p = \hat{p}$.

  When $0 \leq p < \hat{p}$, we note that because $(\beta x - \gamma)x$ is parabolic with a minimum at $x = \gamma/2\beta$, and $\hat{S}_{-1} \geq \frac{\gamma}{2\beta}$, it follows that

$$\nu(p) = \frac{\phi_{\text{inc}}}{\phi_{\text{inc}} + \beta}(\beta\hat{S}_{-1} - \gamma)\hat{S}_{-1} \tag{2.124}$$

  is increasing in $\hat{S}_{-1}$, so $\max_{p\in[0,\hat{p}]} \nu(p)$ is attained on this interval when $\hat{S}_{-1}$ is maximized. Because $-\frac{\gamma}{2\beta}\, W_{-1}(x)$ is monotonically increasing, it follows that $\max_{p\in[0,\hat{p}]} \nu(p)$ is maximal when

$$-2\frac{\beta}{\gamma e}(1-p)(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}(1-p)}$$

  is maximal. Consequently, we need to maximize $-axe^{-x}$ (with $a > 0$), with $x(p) = \frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}(1-p)$, over the interval $0 \leq p \leq \hat{p}$. This corresponds to maximizing $-axe^{-x}$ over $[\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}, \frac{\beta+\phi_{\text{inc}}\alpha}{(1-2\alpha)\beta - \alpha\phi_{\text{inc}}}] \subset [1, \infty)$. Observe that $-axe^{-x}$ has a unique global minimum at $x = 1$, and in particular, it is increasing when $x \geq 1$. This implies that in the relevant range of x, $\hat{S}_{-1}$ increases with x, and thus decreases in p. It follows that $\hat{S}_{-1}$ is maximal when $p = 0$, and its value is

$$\hat{S}_{-1}|_{p=0} = -\frac{\gamma}{2\beta}\, W_{-1}\left(-2\frac{\beta}{\gamma e}(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}}\right). \tag{2.125}$$

  Thus, $\max_{p\in[0,1)} \nu(p) = \nu(0)$, and is attained when $p = 0$ (note also that $\nu(p)$ is continuous at $p = \hat{p}$).

- When $0 < (1-2\alpha)\beta - \alpha\phi_{\text{inc}} \leq \gamma$, $(1-p)((1-2\alpha)\beta - \alpha\phi_{\text{inc}}) \leq \gamma$ is equivalent to $\hat{p} = 1 - \frac{\gamma}{(1-2\alpha)\beta - \alpha\phi_{\text{inc}}} \leq p$, which is satisfied for all $p \in [0,1)$, since $\hat{p} \leq 0$. Thus, $\nu(p)$ is a decreasing function of p for all $p \in [0,1)$, and thus $\max_{p\in[0,1)} \nu(p) = \phi_{\text{inc}}(1-\alpha)\alpha$ is attained at $p = 0$.

- When $\alpha\phi_{\text{inc}} \geq (1-2\alpha)\beta$, then $(1-p)((1-2\alpha)\beta - \alpha\phi_{\text{inc}}) \leq \gamma$ is always satisfied (since the left hand side is never positive, and $\gamma > 0$). Thus, $\nu(p) = \phi_{\text{inc}}(1-p)^2(1-\alpha)\alpha$, which decreases with p, so $\max_{p\in[0,1)} \nu(p) = \phi_{\text{inc}}(1-\alpha)\alpha$, and is attained at $p = 0$.

Rearranging the conclusions of the preceeding discussion, we see that

$$\max_{p\in[0,1),t\geq 0} \dot{V} = \begin{cases} \frac{\phi_{\text{inc}}}{\phi_{\text{inc}}+\beta}(\beta\hat{S}_{-1} - \gamma)\hat{S}_{-1}|_{p=0} & \text{if } 0 \leq \alpha\phi_{\text{inc}} < (1-2\alpha)\beta - \gamma\,, \\ \alpha\phi_{\text{inc}}(1-\alpha) & \text{if } (1-2\alpha)\beta - \gamma \leq \phi_{\text{inc}}\alpha\,. \end{cases} \tag{2.126}$$

When $0 \leq \alpha\phi_{\text{inc}} < (1-2\alpha)\beta - \gamma$, then $\max_{p\in[0,1),t\geq 0} \dot{V} = \frac{\phi_{\text{inc}}}{\phi_{\text{inc}}+\beta}(\beta\hat{S}_{-1} - \gamma)\hat{S}_{-1}|_{p=0}$, which increases as $S_{-1}|_{p=0}$ increases, as stated earlier. Since $-W_{-1}(x)$ increases with x, and $-2\frac{\beta}{\gamma e}(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}}$ is an increasing function of ϕ_{inc}, we conclude that in this range, $\max_{p\in[0,1),t\geq 0} \dot{V}$ increases with ϕ_{inc}. When $(1-2\alpha)\beta - \gamma \leq \phi_{\text{inc}}\alpha$, $\max_{p\in[0,1)} \nu(p)$ manifestly increases linearly with ϕ_{inc}. In all, $\max_{p\in[0,1),t\geq 0} \dot{V}$ is a monotonically increasing function of ϕ_{inc}.

Note that at the point separating the two regimes, $\phi_{\text{inc}} = \frac{(1-2\alpha)\beta-\gamma}{\alpha} = 16570.71/\text{day}$, $\max_{p\in[0,1),t\geq 0} \dot{V} = \big((1-2\alpha)\beta - \gamma\big)(1-\alpha) = 0.29/\text{day}$ (with parameters as in table 2.1).

Finally, to obtain a value of ϕ_{inc} that yields a maximal vaccination rate of $\Phi = 0.1/\text{day}$ (as was estimated in [37]), we solve

$$\frac{\phi_{\text{inc}}}{\phi_{\text{inc}}+\beta}(\beta\hat{S}_{-1} - \gamma)\hat{S}_{-1}|_{p=0} = \Phi\,, \tag{2.127}$$

which gives $\hat{S}_{-1}|_{p=0} = \frac{\gamma \pm \sqrt{\gamma^2 + 4\beta\Phi(1+\beta/\phi_{\text{inc}})}}{2\beta}$. We take the solution corresponding to the positive sign (the other one gives negative $\hat{S}_{-1}$, which is biologically absurd). Thus,

$$W_{-1}\left(-2\frac{\beta}{\gamma e}(1-\alpha)e^{-\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma}}\right) = -1 - \sqrt{1 + 4\frac{\beta}{\gamma^2}\Phi(1+\beta/\phi_{\text{inc}})}\,, \tag{2.128}$$

which is equivalent to

$$\frac{2\beta(1-\alpha)}{\gamma\big(1+\sqrt{1+4\frac{\beta}{\gamma^2}\Phi(1+\beta/\phi_{\text{inc}})}\big)} = \exp\left(\frac{\beta+\phi_{\text{inc}}\alpha}{\gamma} - \sqrt{1+4\frac{\beta}{\gamma^2}\Phi(1+\beta/\phi_{\text{inc}})}\right),$$

which we solve numerically for ϕ_{inc}, with parameters as in Tables 2.1 and 2.2, to get $\phi_{\text{inc}} \approx 5190/\text{day}$.

2.G The individual equilibrium

In this section, we show that for each of the five models defined in § 2.6, the game defined in § 2.4 always has a unique convergently stable Nash equilibrium (defined in § 2.4 and abbreviated CSNE). The proofs given here are constructive, *i.e.,* they also provide a method for numerically finding the individual equilbrium (p_{i}).

2.G.1 Vaccination rate $\propto$ disease prevalence

In this scenario we have 3 cases to examine:

1. $p_m \geq 1$ ($\iff \alpha\phi_{\text{prev}} \geq \gamma(1-\alpha)$)
2. $p_m \leq 1$ and $p_0 \leq 0$
3. $p_m \leq 1$ and $p_0 \in (0,1)$

Recall that in the first case, we have proven (in appendix 2.E.1) that π_p decreases with p. As stated in §2.6.1, we assume that π_p decreases with p for the other two cases as well.

$p_m \geq 1$

In this case, we have a unique convergently stable Nash equilibrium (CSNE). To see this, note that $\psi_p + \pi_p = 1$ and $\partial_p \pi_p < 0$. Thus,

$$\begin{aligned}\Delta E &= [\pi_p + (1-\pi_p)r - r/a]a(P-Q) \\ &= [\pi_p a(1-r) - r(1-a)](P-Q) \\ &= a(1-r)\left[\pi_p - \frac{r(1-a)}{a(1-r)}\right](P-Q)\,. \end{aligned} \tag{2.129}$$

It is convenient to define

$$\rho_1 = \frac{r(1-a)}{a(1-r)} = \frac{r}{1-r}\Big/\frac{a}{1-a}\,, \tag{2.130}$$

which we can interpret as an ***odds ratio***, namely the odds of a bad outcome from vaccination (compared with infection) relative to the odds of an outbreak occuring. The odds ratio is well-defined and strictly positive ($\rho_1 > 0$) because $0 < r < 1$ and $0 < a < 1$. Since π_p decreases monotonically with p, there are three cases:

- If $\pi_0 \leq \rho_1$ then $\pi_p < \rho_1 \quad \forall p > 0$. It follows that $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P \iff P = 0$. Hence $p_{\text{i}} = 0$ is the unique Nash equilibrium. Let $0 \leq P < Q$ and fix $\epsilon \in [0,1)$. It follows that $p > 0$, and so $\pi_p - \rho_1 < 0$. Thus $\Delta E > 0$ and p_{i} is convergently stable.
- If $\alpha = \pi_1 \geq \rho_1$, then $\pi_p > \rho_1 \quad \forall p < 1$. It follows that $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P \iff P = 1$. Hence $p_{\text{i}} = 1$ is the unique Nash equilibrium. The condition translates to $r(1-a) < \alpha a(1-r)$, or $r_{\text{v}} < a r_{\text{i}}\alpha + a r_{\text{v}}(1-\alpha)$. Recall that if an outbreak occurs, at the end of the epidemic individuals have either been vaccinated or have contracted the disease. Thus the right hand side is the risk to a vaccinator, and the left hand side is the minimal possible risk to a delayer (assuming no-one

is infected after the initial outbreak; if there are secondary infections, then because $r_{\text{v}} < r_{\text{i}}$, the delayer's risk can only be increased). Let $1 \geq P > Q$ and fix $\epsilon \in [0, 1)$. It follows that $p < 1$, and so $\pi_p - \rho_1 > 0$. Thus $\Delta E > 0$ and p_{i} is convergently stable.

- If $\pi_0 > \rho_1 > \pi_1 = \alpha$ then there is a unique $\tilde{p} \in (0, 1)$ such that $\pi_p - \rho_1 > 0$ if $p < \tilde{p}$, $\pi_{\tilde{p}} = \rho_1$ and $\pi_p - \rho_1 < 0$ if $p > \tilde{p}$. Now since for any $\epsilon \in [0, 1)$, $Q < P \implies p < P$ and $Q > P \implies p > P$, we have $\forall \epsilon \in [0, 1) \quad \Delta E > 0 \ \forall Q \neq P \iff P = \tilde{p}$ (for other P take Q between P and $\tilde{p}$). Thus, the unique Nash equilibrium p_{i} is the unique solution to $\pi_{p_{\text{i}}} = \rho_1$. Fix $\epsilon \in [0, 1)$. If $Q < P \leq p_{\text{i}}$, $Q \leq p < P \leq p_{\text{i}}$. Thus $\pi_p - \rho_1 > 0 \implies \Delta E > 0$. Similarly, If $Q > P \geq p_{\text{i}}$, $Q \geq p > P \geq p_{\text{i}}$. Thus $\pi_p - \rho_1 < 0 \implies \Delta E > 0$. Hence p_{i} is convergently stable.

  Now, to find $\tilde{p}$: recall that $1 - \pi_p = \frac{\phi_{\text{prev}}}{\beta(1-p)} \ln\left(\frac{\beta}{\phi_{\text{prev}}}(1-\alpha)(1-p) + 1\right)$ and thus

$$\tilde{p} = 1 + \phi_{\text{prev}} \frac{(1-\rho_1) + (1-\alpha)\, W\left(-\frac{1-\rho_1}{1-\alpha} e^{-\frac{1-\rho_1}{1-\alpha}}\right)}{(1-\alpha)\beta(1-\rho_1)}\,. \tag{2.131}$$

  Again, W is applied to a negative argument, and it is necessary to determine which branch of W to use. The principal branch gives $\tilde{p} = 1$, and $\pi_p \to \alpha$ as $p \to 1$, and $\rho_1 > \alpha$ by assumption, and so by elimination we must use W_{-1}. Interestingly, $\tilde{p}$ depends linearly on ϕ_{prev}. Recall that $W_{-1} \leq -1$, and so $\frac{\partial \tilde{p}}{\partial \phi_{\text{prev}}} < 0$.

Thus, in all three cases there exists a unique CNSE.

$p_m \leq 1$ **and** $p_0 \leq 0$

In the 2nd case, recall that $\psi_p = \frac{\phi_{\text{prev}}}{\gamma}\pi_p$, and let

$$\rho_2 = \frac{r}{a(1 + r\phi_{\text{prev}}/\gamma)}\,, \tag{2.132}$$

to obtain

$$\begin{aligned}
\Delta E &= [\pi_p + \frac{\phi_{\text{prev}}}{\gamma}\pi_p r - r/a]a(P-Q) \\
&= [\pi_p a(1 + \frac{\phi_{\text{prev}}}{\gamma} r) - r](P-Q) \\
&= a(1 + \frac{\phi_{\text{prev}}}{\gamma} r)[\pi_p - \frac{\gamma r}{a(r\phi_{\text{prev}} + \gamma)}](P-Q) \\
&= a(1 + \frac{\phi_{\text{prev}}}{\gamma} r)[\pi_p - \rho_2](P-Q)\,.
\end{aligned} \tag{2.133}$$

Now recall that π_p decreases with p. So, an identical argument to the one in appendix 2.G.1 also applies here:

- If $\pi_0 \leq \rho_2$ then $\pi_p < \rho_2 \quad \forall p > 0$. Thus, $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P \iff P = 0$. Hence $p_{\text{i}} = 0$ is the unique Nash equilibrium. Let $0 \leq Q < P < 1$ and fix $\epsilon \in [0,1)$. It follows that $p > 0$, and so $\pi_p - \rho_2 < 0$. Thus $\Delta E > 0$ and p_{i} is convergently stable.

- If $\alpha = \pi_1 \geq \rho_2 \geq 0$, then $\pi_p > \rho_2 \quad \forall p < 1$. It follows that $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P \iff P = 1$. Hence $p_{\text{i}} = 1$ is the unique Nash equilibrium. Let $1 \geq P > Q$ and fix $\epsilon \in [0,1)$. It follows that $p < 1$, and so $\pi_p - \rho_2 > 0$. Thus $\Delta E > 0$ and p_{i} is convergently stable.

- If $\pi_0 > \rho_2 > \pi_1 = \alpha$ then there is a unique $\tilde{p} \in (0,1)$ such that $\pi_p - \rho_2 > 0$ if $p < \tilde{p}$, $\pi_{\tilde{p}} = \rho_2$ and $\pi_p - \rho_2 < 0$ if $p > \tilde{p}$. Now since for any $\epsilon \in [0,1)$, $Q < P \implies p < P$ and $Q > P \implies p > P$, we have $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P \iff P = \tilde{p}$. Thus, the unique Nash equilibrium p_{i} is the unique solution to $\pi_{p_{\text{i}}} = \rho_2$. Fix $\epsilon \in [0,1)$. If $Q < P \leq p_{\text{i}}$, $Q \leq p < P \leq p_{\text{i}}$. Thus $\pi_p - \rho_2 > 0 \implies \Delta E > 0$. Similarly, If $Q > P \geq p_{\text{i}}$, $Q \geq p > P \geq p_{\text{i}}$. Thus $\pi_p - \rho_2 < 0 \implies \Delta E > 0$. Hence p_{i} is convergently stable.

  To find $\tilde{p}$, recall that $R^1_\infty = (1-p)\pi_p$. Furthermore, from equation (2.55), we have

$$
\begin{aligned}
R^1_\infty &= \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\beta S^1_\infty + \phi_{\text{prev}}}\right) \\
S^1_\infty &= 1 - p - \frac{\phi_{\text{prev}} + \gamma}{\gamma} R^1_\infty \\
&\Downarrow \\
R^1_\infty &= \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\beta(1 - p - \frac{\phi_{\text{prev}}+\gamma}{\gamma} R^1_\infty) + \phi_{\text{prev}}}\right) \\
&\Downarrow \\
(1-p)\pi_p &= \frac{\gamma}{\beta} \ln\left(\frac{\beta S(0) + \phi_{\text{prev}}}{\beta(1-p)\left(1 - \frac{\phi_{\text{prev}}+\gamma}{\gamma}\pi_p\right) + \phi_{\text{prev}}}\right) .
\end{aligned}
$$

  Substituting $p = \tilde{p}$, and using $\pi_{\tilde{p}} = \rho_2 = \frac{\gamma r}{a(r\phi_{\text{prev}}+\gamma)}$, we obtain after minor rearrangement

$$
\frac{\beta(1-\tilde{p})r}{a(r\phi_{\text{prev}} + \gamma)} = \ln\left(\frac{a(\beta(1-\tilde{p})(1-\alpha) + \phi_{\text{prev}})(r\phi_{\text{prev}} + \gamma)}{\beta(1-\tilde{p})(\gamma(a-r) - (1-a)r\phi_{\text{prev}}) + \phi_{\text{prev}} a(r\phi_{\text{prev}} + \gamma)}\right) . \tag{2.134}
$$

  However, we have not succeeded in obtaining an analytical solution for the individual

equilibrium from this equation.

Thus, in all three cases there exists a unique CSNE.

$p_m \leq 1$ **and** $p_0 \in (0,1)$

Since π_p decreases with p, the argument above shows that there is a unique CSNE in each of the the two intervals $[0, p_0]$ and $[p_0, 1]$, denoted $P_{I,1}$ and $P_{I,2}$, respectively. These are the only candidates for Nash equilibria in the interval $[0,1]$: Adding the two sub-intervals together amounts to adding more strategies to the game. Thus, a strategy which was a Nash equilibrium in one of the sub-intervals may not be a Nash equilibrium for the larger strategy set (because players now have a larger strategy set to choose from). However, a strategy which is a Nash equilibrium for $[0,1]$ must be a Nash equilibrium in any sub-interval of $[0,1]$ which contains it. The situation for convergent stability is a bit more subtle, and is considered below.

Note that when $p \neq p_0$,

$$\operatorname{sign}(\Delta E) = \operatorname{sign}\big((\pi_p - \rho(p))(P - Q)\big) \tag{2.135}$$

where

$$\rho(p) = \begin{cases} \rho_1 & \text{if } p < p_0, \\ \rho_2 & \text{if } p \geq p_0\,. \end{cases} \tag{2.136}$$

Note also that

$$\pi_{p_0} = \rho_1 \iff r = \frac{\gamma a}{\phi_{\text{prev}}(1-a) + \gamma} \iff \pi_{p_0} = \rho_2\,. \tag{2.137}$$

This may seem slightly perplexing at first, but recall that $\pi_{p_0} = \frac{\gamma}{\phi_{\text{prev}}+\gamma}$. Thus, if it so happens that $\pi_{p_0} = \rho_1$ or $\pi_{p_0} = \rho_2$, r and a must be related so that in fact $\rho_1 = \rho_2$.

We must now check a number of cases:

1. $\pi_0 \leq \rho_1$ and $\pi_{p_0} < \rho_2$: In this case, $\pi_p < \rho(p) \quad \forall p \in (0,1]$, and $P_{I,1} = 0$ and $P_{I,2} = p_0$. But if $P = p_0$ and $Q \in [0, p_0)$, $\Delta E < 0$ and so $P = p_0$ cannot be a Nash equilibrium. However, if $P = 0$, then $\forall \epsilon \in [0,1) \quad \Delta E > 0 \; \forall Q \neq P$, and so $P = 0$ is a Nash equilibrium. This is trivial for $Q \leq p_0$. For $p_0 < Q$, we have $\pi_p - \rho(p) < 0$ and $P - Q < 0$, so $\Delta E > 0$ as required. For convergent stability, we only need to check that if $0 < P < p_0 < Q \leq 1$, then for any $\epsilon \in [0,1)$ we have $\Delta E > 0$. In this case, $p \in [P, Q]$, and $P - Q < 0$. Furthermore, $\pi_p - \rho_1 < 0$ for any $p \leq p_0$, and $\pi_p - \rho_2 < 0$ for any $p \geq p_0$. But since $\pi_{p_0} - \rho_2 < 0$, $\pi_{p_0} - \rho_1 = 0$ is impossible and from continuity, $\pi_p - \rho_1 < 0$. Thus, $\Delta E(P, Q, \epsilon) > 0$ as required for convergent stability.

2. $\pi_0 \leq \rho_1$ and $\pi_{p_0} > \rho_2 > \pi_1 = \alpha$: In this case, $P_{I,1} = 0$ and $P_{I,2} \in (p_0, 1)$. Since $P_{I,1} = 0$ is a CSNE in $[0, p_0]$, we know that for any $\epsilon \in [0, 1)$ and $0 < P < Q \leq p_0$, $\Delta E > 0$. In particular, for $\epsilon = 0$, and $0 < P < Q = p_0$ we get $[\pi_{p_0} + \psi_{p_0} r - r/a] = \frac{\Delta E}{(P-Q)} < 0$. But, from $P_{I,2} \in (p_0, 1)$ we can similarly get (for $\epsilon = 0$, $p_0 = Q < P < P_{I,2}$) $[\pi_{p_0} + \psi_{p_0} r - r/a]a = \frac{\Delta E}{(P-Q)} > 0$, a contradiction.

3. $\pi_0 \leq \rho_1$ and $\alpha = \pi_1 \geq \rho_2 > 0$: Here, $P_{I,1} = 0$ and $P_{I,2} = 1$. Since $P_{I,1} = 0$ is a CSNE in $[0, p_0]$, $\forall \epsilon \in [0, 1)$ and $0 \leq P < Q \leq p_0$, $\Delta E > 0$. In particular, for $\epsilon = 0$, and $0 < P < Q = p_0$ we get $[\pi_{p_0} + \psi_{p_0} r - r/a] = \frac{\Delta E}{(P-Q)} < 0$. Similarly, since $P_{I,2} = 1$ is a CSNE in $[p_0, 1]$, $\forall \epsilon \in [0, 1)$ and $p_0 \leq Q < P \leq 1$, $\Delta E > 0$. In particular, for $\epsilon = 0$, and $p_0 = Q < P \leq 1$ we get $[\pi_{p_0} + \psi_{p_0} r - r/a] = \frac{\Delta E}{(P-Q)} > 0$, which is a contradiction.

4. $\rho_2 \geq \pi_{p_0} \geq \rho_1$: In this case, simple algebra gives $\rho_1 = \rho_2 = \pi_{p_0}$ and $P_{I,1} = p_0$ and $P_{I,2} = p_0$. Thus, it follows that p_0 is the unique CSNE in the interval $[0, 1]$.

5. $\pi_{p_0} > \rho_1$ and $\alpha = \pi_1 \geq \rho_2$: In a manner analogous to the first case, here, $p_{\text{i}} = 1$ is the unique CSNE.

6. $\pi_{p_0} > \rho_1$ and $\pi_{p_0} > \rho_2 > \pi_1 = \alpha$: , $P_{I,1} = p_0$ and $P_{I,2} \in (p_0, 1)$. $P_{I,1} = p_0$ cannot be a Nash equilibrium since for $p_0 < Q < P_{I,2}$, and any $\epsilon \in [0, 1)$, $\Delta E < 0$ since $P_{I,2}$ is the unique CSNE in $[p_0, 1]$. To show that $P_{I,2}$ is a Nash equilibrium, fix $\epsilon \in [0, 1)$ and $P = P_{I,2}$ and let $Q < P_{I,2}$. Note that $\pi_p > \rho_1 \ \forall p \in [0, p_0]$ and that because π_p is decreasing and $\pi_{P_{I,2}} = \rho_2$, $\pi_p > \rho_2 \ \forall p < P_{I,2}$. Thus $\pi_p > \rho(p) \ \forall p < P_{I,2}$ and so in particular, $\Delta E > 0$ and $P_{I,2}$ is the unique Nash equilibrium. To see that $P_{I,2}$ is also convergently stable, we must only show that for any P and Q, $0 \geq Q < p_0 < P < P_{I,2}$ and $\epsilon \in [0, 1)$, we have $\Delta E > 0$. But under these conditions, $p < P_{I,2}$ and so again $\pi_p > \rho(p)$, which implies $\Delta E > 0$, as required.

7. $\pi_0 > \rho_1 > \pi_{p_0}$ and $\pi_{p_0} < \rho_2$: Now, $P_{I,1} \in (0, p_0)$ and $P_{I,2} = p_0$. Similar to the above case, this implies that $P_{I,1}$ is the unique CSNE (given by equation (2.131)).

8. $\pi_0 > \rho_1 > \pi_{p_0}$ and $\alpha = \pi_1 \geq \rho_2$: Simple algebra shows that this case is impossible:

$$\begin{aligned} \rho_1 > \pi_{p_0} &= \frac{\gamma}{\gamma + \phi_{\text{prev}}} \\ r(1-a)(\gamma + \phi_{\text{prev}}) &> a(1-r)\gamma \\ r((1-a)\phi_{\text{prev}} + \gamma) &> a\gamma \end{aligned}$$

but

$$\begin{aligned} \rho_2 \leq \pi_1 < \pi_{p_0} = \frac{\gamma}{\gamma + \phi_{\text{prev}}} \\ r(\gamma + \phi_{\text{prev}}) < a(\phi_{\text{prev}} r + \gamma) \\ r((1-a)\phi_{\text{prev}} + \gamma) < a\gamma \,. \end{aligned}$$

9. $\pi_0 > \rho_1 > \pi_{p_0}$ and $\pi_{p_0} > \rho_2 > \pi_1$: The reasoning applied to show that the case above is impossible also rules this case out.

We conclude that in all cases there is a unique CSNE, which we denote by p_{i}.

2.G.2 Vaccination rate $\propto$ incidence

Since $\partial_p \psi_p < 0$ and $\partial_p \pi_p < 0$, $\partial_p (\pi_p + r\psi_p) < 0$. Thus, an identical argument to the one given in appendix 2.G.3 allows us to show that there is always a CSNE for this model. In particular, there are three possibilities:

- If $\pi_0 + r\psi_0 \leq r/a$ then $p_{\text{i}} = 0$ is a unique CSNE.
- If $\alpha = \pi_1 + r\psi_1 \geq r/a$, then $p_{\text{i}} = 1$ is a unique CSNE.
- If $\pi_0 + r\psi_0 > r/a > \pi_1 + r\psi_1 = \alpha$ then there is a unique CSNE, $p_{\text{i}} \in (0,1)$ such that $\pi_{p_{\text{i}}} + r\psi_{p_{\text{i}}} = r/a$. To simplify this last condition, we use equations (2.84) to obtain

$$0 = \pi_{p_{\text{i}}} + \psi_{p_{\text{i}}} r - \frac{r}{a} = \frac{a\left(\beta + r(1-\alpha)\phi_{\text{inc}}\right) - r(\beta + \phi_{\text{inc}})}{a(\beta + \phi_{\text{inc}})} - \frac{\beta + r\phi_{\text{inc}}}{\beta + \phi_{\text{inc}}} \frac{S_\infty}{1 - p_{\text{i}}}$$

which is equivalent to

$$\frac{S_\infty}{1 - p_{\text{i}}} = \frac{a\left(\beta + \alpha\phi_{\text{inc}} + r(1-\alpha)\phi_{\text{inc}}\right) - r(\beta + \phi_{\text{inc}})}{a(\beta + r\phi_{\text{inc}})} \,. \tag{2.138}$$

Plugging equation (2.138) into equation (2.78) and rearranging gives the individual equilibrium,

$$\begin{aligned} p_{\text{i}} = 1 + \frac{a\gamma(\beta + r\phi_{\text{inc}})}{r(\beta + \phi_{\text{inc}})(\beta + a\alpha\phi_{\text{inc}})} \\ \times \ln\left(\frac{a\left(\beta + \alpha\phi_{\text{inc}} + r(1-\alpha)\phi_{\text{inc}}\right) - r(\beta + \phi_{\text{inc}})}{a(1-\alpha)(\beta + r\phi_{\text{inc}})}\right) \,. \end{aligned} \tag{2.139}$$

2.G.3 Vaccination rate $\propto$ proportion still susceptible

Recall that for this model, $\psi_p = 1-\pi_p$ (see appendix 2.E.3) and that we assume π_p decreases with p (as stated in § 2.6.3). Thus, $\partial_p\left(\pi_p + r\psi_p\right) = \partial_p\left((1-r)\pi_p\right) < 0$, and we can infer that:

- If $\pi_0 + r\psi_0 \leq r/a$ then $\pi_p + r\psi_p < r/a \quad \forall p > 0$. Hence $p_{\text{i}} = 0$ is the unique Nash equilibrium. From reasoning similar to that given in appendix 2.G.1, p_{i} is convergently stable.
- If $\alpha = \pi_1 + r\psi_1 \geq r/a$, then $\pi_p + r\psi_p > r/a \quad \forall p < 1$. It follows that $p_{\text{i}} = 1$ is the unique Nash equilibrum and that it is convergently stable. The condition is equivalent to $a\alpha r_{\text{i}} \geq r_{\text{v}}$, which is quite intuitive: $a\alpha r_{\text{i}}$ is a lower bound on the cost for a delayer (attained when there are no new cases after the initial outbreak). If even this lower bound is higher than the cost of vaccinating, then clearly everyone should vaccinate.
- If $\pi_0 + r\psi_0 > r/a > \pi_1 + r\psi_1 = \alpha$ then there is a unique CSNE, $p_{\text{i}} \in (0,1)$ such that $\pi_{p_{\text{i}}} + r\psi_{p_{\text{i}}} = r/a$.

Thus, there is always a CSNE for this model.

2.G.4 Instantaneous vaccination of a proportion ϕ_{inst} of the population

In this case, $\psi_p = \phi_{\text{inst}}(1-\alpha)$ and $\pi_p = 1 - \psi_p - \frac{S_\infty}{1-p}$. Thus,

$$\Delta E = [\pi_p + \phi_{\text{inst}}(1-\alpha)r - r/a]a(P-Q)\,. \tag{2.140}$$

Letting $\rho = r/a - r\phi_{\text{inst}}(1-\alpha)$, since π_p decreases monotonically with p, we can use an argument similar to those used for the models considered above to show that:

- if $\pi_0 \leq \rho$ the $p_{\text{i}} = 0$ is the unique CSNE.
- if $\pi_1 \geq \rho$ then $\pi_p > \rho$ for any $p \in [0,1)$ and so $p_{\text{i}} = 1$ is the unique CSNE. Rearranging the condition $\pi_1 \geq \rho$ gives $a\alpha + ra\phi_{\text{inst}}(1-\alpha) \geq r$. This admits a simple biological interpretation: $a\alpha + ra\phi_{\text{inst}}(1-\alpha)$ is the relative risk of delaying when the epidemic does not successfully spread (that is, one can only be infected during the initial outbreak). Thus, if the risks of delaying are greater than vaccinating even if the disease does not spread beyond the cohort initially infected in the outbreak, then it is worthwhile for individuals to vaccinate pre-emptively.
- if $\pi_0 > \rho > 0$ then there is a unique CSNE, $p_{\text{i}} \in (0,1)$ such that $\pi_{p_{\text{i}}} = \rho$. In order to find p_{i} explicitly, we use equation (2.110) and substitute $\pi_{p_{\text{i}}} = r/a - r\phi_{\text{inst}}(1-\alpha)$:

$$r/a - r\phi_{\text{inst}}(1-\alpha) = -\frac{\gamma}{\beta(1-p_{\text{i}})}\ln\left(\frac{1 - r/a - (1-r)\phi_{\text{inst}}(1-\alpha)}{(1-\alpha)(1-\phi_{\text{inst}})}\right)\,,$$

and consequently,

$$p_{\text{i}} = 1 + \frac{a\gamma}{r\beta(1 - a\phi_{\text{inst}}(1-\alpha))} \ln\left(\frac{a\left(1-(1-r)\phi_{\text{inst}}(1-\alpha)\right) - r}{a(1-\alpha)(1-\phi_{\text{inst}})}\right). \quad (2.141)$$

2.G.5 constant rate vaccination

As in [37], we assume π_p is a decreasing function of p. The analysis (performed originally in [37]) is then identical to appendix 2.G.1, implying the existence of a unique CSNE, which we denote by p_{i}. Using the definition of ρ_1 given in equation (2.130), we have

- if $\pi_1 \geq \rho_1$ then $p_{\text{i}} = 1$.
- if $\pi_1 < \rho_1 < \pi_0$ then p_{i} is the unique solution of $\pi_{p_{\text{i}}} = \rho_1$
- if $\rho_1 \geq \pi_0$ then $p_{\text{i}} = 0$.

2.H The group optimum

We have obtained an analytical formula for the group optimum (defined in § 2.5), for one sub-case of one of our models. The calculation is given below.

2.H.1 Vaccination rate $\propto$ disease prevalence

We consider the case when $p_m \geq 1$. Recall that if $p_m \geq 1$, then $S_\infty = 0$, $\psi_p = \frac{\phi_{\text{prev}}}{\beta(1-p)} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right)$, and $\pi_p = 1 - \psi_p$ (see appendix 2.E.1). Thus,

$$\begin{aligned}
C(p) &= rp + (1-p)a(1-(1-r)\psi_p) \\
&= rp + (1-p)a\left(1 - (1-r)\frac{\phi_{\text{prev}}}{\beta(1-p)} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right)\right) \\
&= rp + a\left((1-p) - (1-r)\frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}} S(0) + 1\right)\right) \\
C(0) &= a\left(1 - (1-r)\frac{\phi_{\text{prev}}}{\beta} \ln\left(\frac{\beta}{\phi_{\text{prev}}}(1-\alpha) + 1\right)\right) \\
C(1) &= r \\
C'(p) &= r + a\left(-1 - (1-r)\frac{\phi_{\text{prev}}}{\beta}\left(-\frac{\beta}{\phi_{\text{prev}}}(1-\alpha)\right)\frac{1}{\frac{\beta}{\phi_{\text{prev}}} S(0) + 1}\right) \\
&= r + a\left(\frac{\phi_{\text{prev}}(1-\alpha)(1-r)}{\beta S(0) + \phi_{\text{prev}}} - 1\right)
\end{aligned} \tag{2.142}$$

Note that $C'(p)$ increases with p since $S(0)$ decreases with p and critical points of $C(p)$ are minima. Thus, if there is a critical point within $[0,1]$, then it is the global minimum; otherwise, the global minimum is at $C(0)$. To find critical points, set $C'(p) = 0$ or equivalently,

$$\frac{r}{a} = 1 - \frac{\phi_{\text{prev}}(1-\alpha)(1-r)}{\beta S(0) + \phi_{\text{prev}}} \tag{2.143}$$

which can only happen if $r < a$ (since $\alpha < 1$ and $r < 1$), that is, the relative risk (of vaccination *versus* infection) is less than the probability of an outbreak. If $r \geq a$ then $C'(p) \geq 0$ throughout $[0,1]$ and the minimal group cost $C(p)$ is attained at $p_{\text{g}} = 0$. This is easily explained: If $r_{\text{v}} \geq a r_{\text{i}}$ then the mortality risks from vaccination are no less than those of dying in an outbreak. In this case, vaccinating is not worthwhile for either the individual or the group. We now solve equation (2.143) for the initial coverage p at the critical point of the group cost, $C(p)$, assuming $r < a$:

$$\begin{aligned}
\beta S(0) &= \frac{\phi_{\text{prev}}(1-\alpha)a(1-r)}{a-r} - \phi_{\text{prev}} \\
1 - p &= \frac{\phi_{\text{prev}}}{\beta(1-\alpha)}\left(\frac{(1-\alpha)a(1-r)}{a-r} - 1\right) \\
p &= 1 - \frac{\phi_{\text{prev}}}{\beta(1-\alpha)}\left(\frac{(1-a)r - (1-r)\alpha a}{a-r}\right)
\end{aligned} \tag{2.144}$$

The critical point is attained at $p \geq 1$ if and only if $(1-a)r \leq (1-r)\alpha a$, which is equivalent to

$$r_{\mathrm{v}} \leq a(r_{\mathrm{i}}\alpha + r_{\mathrm{v}}(1-\alpha)) \,, \tag{2.145}$$

and in this case the group optimum is vaccinating the entire population ($p_{\mathrm{g}} = 1$). Biologically, equation (2.145) means that more people are expected to die if, in case of an outbreak, all individuals not infected initially are vaccinated (discounted by the outbreak probability, a), than the number of people expected to die if the entire population is vaccinated pre-emptively. To see this, note the probability of death due to vaccinating is r_{v} (the left hand side of equation (2.145)). To interpret the right hand side of equation (2.145), consider an individual who is not vaccinated pre-emptively. If there is an outbreak (represented by the factor a), the first term in brackets ($r_{\mathrm{i}}\alpha$) represents the probability of being in the initially infected cohort (α), and then dying due to the disease. The second term ($r_{\mathrm{v}}(1-\alpha)$), represents the probability of not being in the initially infected cohort, and dying due to the vaccine side effects. Note that because in this scenario $S_\infty = 0$, no delayers remain susceptible (they are either infected or vaccinated). Thus, any individual who is not pre-emptively vaccinated, and who is not in the initially infected cohort ($1-\alpha$) has either a probability r_{v} of dying due to vaccine side effects, or a probability r_{i} of dying due to the disease. But since $r_{\mathrm{v}} < r_{\mathrm{i}}$, the term $r_{\mathrm{v}}(1-\alpha)$ is a lower bound on the probability of death for an individual who is susceptible immediately after the outbreak is seeded (that is, not pre-emptively vaccinated, or in the cohort initially infected at the beginning of the outbreak).

Note that not vaccinating anyone pre-emptively is the group optimum ($p_{\mathrm{g}} \leq 0$) *iff*

$$\beta(1-\alpha)(a-r) \leq \phi_{\text{prev}} \left((1-a)r - (1-r)\alpha a\right) \,, \tag{2.146}$$

(but this condition is difficult to interpret biologically).

Lastly, if $\phi_{\text{prev}} \leq \beta \frac{(a-r)(1-\alpha)}{r(1-a(1-\alpha))-a\alpha}$ and $(1-a)r > (1-r)\alpha a$, then $p_{\mathrm{g}} \in [0,1)$ and is given by equation (2.144). It is thus interesting to note that p_{g} depends piece-wise linearly on ϕ_{prev}.

Chapter 3

Evolutionary stability in continuous nonlinear public goods games

Chai Molina and David J. D. Earn

3.1 Abstract

We investigate a type of public goods games played in groups of individuals who choose how much to contribute towards the production of a common good, at a cost to themselves. In these games, the common good is produced based on the sum of contributions from all group members, then equally distributed among them. In applications, the dependence of the common good on the total contribution is often nonlinear (*e.g.,* exhibiting synergy or diminishing returns). To date, most theoretical and experimental studies have addressed scenarios in which the set of possible contributions is discrete. However, in many real-world situations, contributions are continuous (*e.g.,* individuals volunteering their time). The "n-player snowdrift games" that we analyze involve continuously varying contributions. We establish under what conditions populations of contributing (or "cooperating") individuals can evolve and persist. Previous work on snowdrift games, using adaptive dynamics, has found that what we term an "equally cooperative" strategy is *locally* convergently and evolutionarily stable. Using static evolutionary game theory, we find conditions under which this strategy is actually *globally* evolutionarily stable. All these results refer to stability to invasion by a single mutant. We broaden the scope of existing stability results by showing that the equally cooperative strategy is locally stable to potentially large population perturbations, *i.e.,* allowing for the possibility that mutants make up a non-negligible proportion of the population (due, for example, to genetic drift, environmental variability or dispersal).

3.2 Introduction

Public goods games [54, 108] arise in a wide variety of biological and social contexts, ranging from microbial evolution [34, 109], tumor growth [110], the evolution of virulence [111] and host manipulation by parasites [112], to cooperative nesting and brood care [113, 114, 115, 116], the evolution of eusociality [117], fisheries management [118] and family economics [119]. These are games played among groups of individuals, who may choose to *cooperate* and contribute towards the production or attainment of a common good at a cost to themselves, or to *defect* and contribute nothing. The common good is then distributed among all members of the group (regardless of whether or not they contributed) [120]. This situation is analogous to Hardin's "Tragedy of the Commons" [121], in which the cost of using a common resource is distributed among group members, but the benefit is personal (*e.g.,* intrabrood competition for parental investment [116]). In both cases, those who act selfishly (by refraining from contribution or by overexploitation), do better than group members who cooperate (either by contributing or by refraining from over-exploiting the common resource). Because cooperative ventures are ubiquitous in nature [122, 123], much research has been devoted to understanding how cooperation can evolve and persist [20, 21, 22, 124]; see recent reviews by Gavrilets [125] and Gokhale and Traulsen [126].

In experimental economics studies of human behaviour, public goods games are typically set up with a linear relationship between the total cost incurred by group members and the benefit they receive [23, 120, 127]. However, in many biological scenarios, the benefit is a nonlinear function of the total cost [21, 54, 112, 128], as there may be a threshold [129, 130], a synergistic effect of contributions [131, 132, 133], diminishing returns [20, 134, 135], or both synergy and diminishing returns [112, 131, 132, 133]. Furthermore, in both theoretical and experimental studies of public goods games, contribution levels are typically taken to be discrete: contribution may be an "all or nothing" affair, whereby a group member can either contribute a fixed, nonzero amount of a resource to the public good, or contribute nothing [20, 128, 129, 131] (usually in studies of the n-player prisoner's dilemma), or, more typically in the economics literature, players are endowed with a number of tokens and decide how many they wish to contribute [23, 120]. In many real situations, however, individuals can vary their degree of cooperation, often continuously [55, 136, 137, 138, 139, 140, 141]. Further realism is often added to models by implementing population structure, [142, 143, 144], but we will avoid this further complication here.

The differences in evolutionary dynamics between 2- and n- player snowdrift games have been studied in games with binary strategies [129, 145]. In public goods games with a continuum of possible contributions, played in unstructured population, studies have investigated how the process by which individual contributions are aggregated affects the possibility of polymorphism [146, 147]. Others have investigated how variability in group size [148] and population dynamics [149] affect the evolutionary outcomes.

Recently, interest in the influence of the functional form of the benefit of contribution on evolutionary dynamics of the snowdrift and other public goods games has increased. Most often, the effect of how the benefit depends on collective investment is investigated in the context of binary strategies (cooperate or defect) [20, 132, 145], sometimes with the addition of population structure (*e.g.,* [150]). However, Deng and Chu [151] have investigated how evolutionary dynamics in continuous public goods games are influenced by nonlinearities in how collective investment is translated to the public good, using specific functional forms (linear, step function or sigmoid). While most other studies investigate stability of a homogeneous population against mutations that are close to the resident strategy, Deng and Chu were interested in stability against invasion by *any* strategy (in line with the original definition of evolutionary stability [152]). They further considered invasion of populations by non-negligible proportions of invaders, using numerical simulations. Chen *et al.* [153] have used simulations to study a similar game played on a spatial lattice using linear cost and two types of sigmoid benefit functions. They found that contributions to the public good are maximized at intermediate values of the steepness and threshold parameters of the sigmoid functions they used.

In this paper, we analyze a class of nonlinear public goods games with continuously varying contributions in unstructured populations and establish under what conditions populations of contributing (or cooperating) individuals can evolve and persist. Examples of public goods games to which our results apply include any in which the dependence of the benefits on the total cost is decelerating or sigmoidal (initially accelerating but eventually decelerating). Most of the specific public goods games considered in the literature fall in this class. Identifying general conditions for the evolution of cooperative strategies and their resistance to invasion is important, because it sheds light on what features of particular biological systems might be responsible for observed evolutionary outcomes. Moreover, since "all models are wrong" [154] (in the sense that no model can take all aspects of reality into account), general results on cooperation lend credibility to the broader application of qualitative conclusions obtained from highly specialized models of particular biological systems. Lastly, general results such as those obtained here can be useful in situations where exact analytical solution of a mathematical model is difficult or impossible.

In most previous studies on nonlinear public goods games with continuous contributions (*e.g.,* [55, 138, 155]), the framework of *adaptive dynamics* [56, 57, 58, 59] has typically been used to analyze and determine evolutionary outcomes. The adaptive dynamics framework assumes an infinite population of a particular phenotype (that is, contribution level) and investigates evolutionary stability by considering a single mutant of a different type and determining whether it can invade the resident population. Because the population of residents is infinite, the effect of the mutant on the average fitness of the resident strategy is negligible.

Here, we compare the predictions of adaptive dynamics with those of *static evolutionary game theory* [28, 59, 156] applied to a general class of nonlinear public goods games

with continuous contributions. Our analysis still considers the limit of an infinite population, but allows mutants to comprise a *finite proportion* of the population; consequently, mutants can affect the average fitness of the resident population (and of other mutants). Our new analysis extends the predictions of adaptive dynamics on evolutionary and convergent stability (§ 3.4) of a cooperative strategy to biologically plausible scenarios in which genetic drift, migration, and/or environmental variability allow a mutant strategy to make up a significant part of the population (even if it is not selected for when initially rare). Our analysis also generalizes the results of [151] (who used Darwinian Dynamics [141]).

In § 3.3, we motivate and construct the class of nonlinear public goods games that we analyze. § 3.4 briefly reviews the two frameworks that we use to analyze these games. We present our results in §4.3 and proofs in §§3.6, 3.7 and 3.8. Finally, in § 3.9, we discuss our results and suggest directions for further developments.

3.3 Class of public goods games

Consider an infinite, well-mixed population of asexual agents. Assume that reproductive fitness is determined by playing a nonlinear public goods game in randomly-assembled groups of $n > 1$ agents. Let $h \geq 0$ be the focal agent's contribution to the public good, and let H denote the mean contribution by the other $n - 1$ agents in the focal agent's group.

Denote the fitness cost and fitness benefit to the focal agent by $c(h, H)$ and $b(h, H)$, respectively. The ***fitness of the focal agent*** is then

$$W(h, H) = b(h, H) - c(h, H)\,, \tag{3.1}$$

where $b(h, H)$ and $c(h, H)$ are non-negative functions of their arguments.

If the cost of contributing is independent of the other group members' contributions, the focal agent's contribution h can be measured in units of the fitness cost of contribution to the public good. Thus, we henceforth assume (with some abuse of notation)

$$c(h, H) = c(h) = h\,. \tag{3.2}$$

The ***total good*** contributed by all members of the group is

$$\eta(h, H) = h + (n - 1)H\,. \tag{3.3}$$

We assume that the resulting benefit to the focal agent is a continuous function of the total good (*i.e.,* the sum of the individual fitness costs),

$$b\,(h, H) = f\big(\eta(h, H)\big)\,. \tag{3.4}$$

We assume that the resulting benefit to the focal agent is a continuous function of the total good, $b(h, H) = f(\eta(h, H))$. Hence, the focal agent's fitness (3.1) is

$$W(h, H) = f\big(\eta(h, H)\big) - h\,, \tag{3.5}$$

which is a continuous function of h and H. Equations (3.3) and (3.5) define a large class of public goods games, namely, continuous n-player snowdrift (or hawk-dove) games [28, 55], in which the public good is fitness (see appendix 3.C). A particular public goods game in this class is specified by choosing the function $f(\eta)$; see figure 3.1.

Biological intuition suggests that there may be a total contribution threshold, $\eta_{\min} > 0$, below which the marginal benefit of contribution does not outweigh its marginal cost. In that case, $W(h, H)$ decreases for all $h < \eta_{\min} - (n-1)H$. If we define $\eta_{\min} = 0$ in the absence of a range of h over which $W(h, H)$ decreases, then no generality is lost by assuming the existence of a threshold $\eta_{\min} \geq 0$. Below we will see that in the situations we consider the focal agent's fitness has a local minimum if $\eta = \eta_{\min}$; we therefore refer to $\eta_{\min}$ as the ***minimizing total good***.

We restrict the class of games we consider slightly by making the biologically sensible assumption that for any level of mean contribution (H) from the non-focal agents, there is a level of focal agent's contribution (h) beyond which its fitness decreases with its contribution; simply put, *the marginal cost of an increase in contribution eventually outweighs its benefit*. In appendix 3.A, we show that this is equivalent to the existence of $\eta_{\max} > 0$ such that $f(\eta) - \eta$ decreases for $\eta \geq \eta_{\max}$. Consequently, the focal agent's fitness $W(h, H)$ decreases with its contribution h when $\eta(h, H) > \eta_{\max}$. In the situations we consider the focal agent's fitness has a local maximum if $\eta = \eta_{\max}$; consequently, we refer to $\eta_{\max}$ as the ***maximizing total good***.

For convenience, we define

$$h_{\min}(H) = \eta_{\min} - (n-1)H\,, \tag{3.6a}$$
$$h_{\max}(H) = \eta_{\max} - (n-1)H\,. \tag{3.6b}$$

For a given mean contribution H by the non-focal agents, $h_{\min}$ and $h_{\max}$ are the levels of contribution required by the focal agent so that the total good is $\eta_{\min} = \eta(h_{\min}(H), H)$ and $\eta_{\max} = \eta(h_{\max}(H), H)$, respectively. Note that $h_{\min}$ and $h_{\max}$ are always well-defined mathematically but they can be negative and hence not biologically meaningful: if the nonfocal group members contribute $(n-1)H > \eta_{\min}$ then $h_{\min}(H) < 0$ and if $(n-1)H > \eta_{\max}$ then $h_{\max}(H) < 0$.

If for any mean non-focal agents' contribution H and focal agent's contribution, h, the marginal costs of contributing outweigh the marginal benefits, then $W(h, H)$ decreases with h regardless of H. Consequently, the unique evolutionarily stable strategy is not to contribute ($h = H = 0$), and it is convergently stable (§ 3.4). In order to avoid this trivial

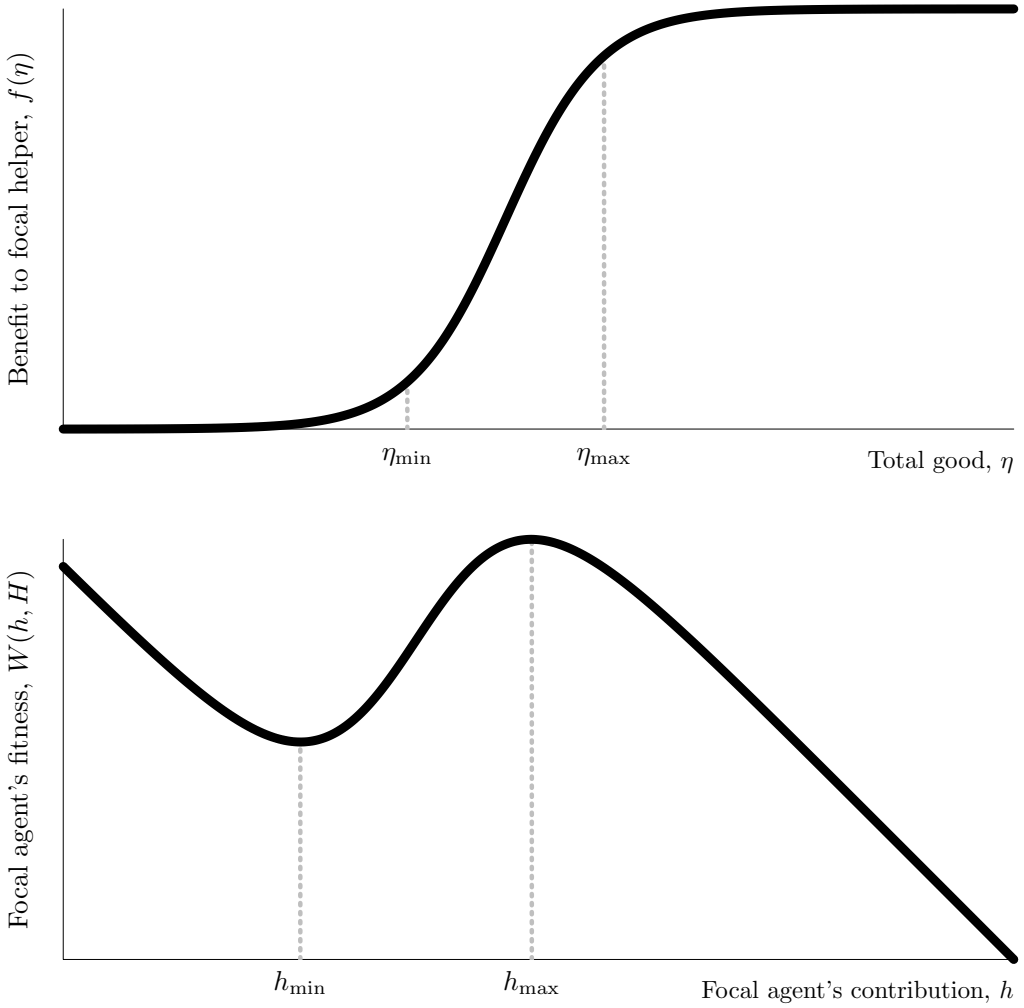

Figure 3.1: Sigmoidal benefit to the focal agent, and its corresponding fitness. Top panel: sigmoidal benefit $f(\eta) = a\big(\beta + \exp(\kappa - b\eta)\big)^{-1} - a\big(\beta + \exp(\kappa)\big)^{-1}$, with $a = 100$, $b = 0.2$, $\beta = 2$, $\kappa = 10$, based on an example from [138] ($\eta_{\min}$ and $\eta_{\max}$ indicated by dashed lines). Bottom panel: The focal agent's fitness $W(h, H)$ (corresponding to the benefit function $f(\eta)$ above) as a function of its contribution h, for fixed mean non-focal agents' contribution $H = 3$ and group size $n = 6$ ($h_{\min}$ and $h_{\max}$ indicated by dashed lines).

outcome, we assume further that

$$\eta_{\min} < \eta_{\max} \,. \tag{3.7}$$

It then follows that fitness $W(h, H)$ decreases with h if the total good $\eta(h, H) < \eta_{\min}$, *i.e.,* for $0 \leq h < \max\{0, h_{\min}(H)\}$. Equivalently, $f(\eta) - \eta$ decreases for $0 \leq \eta < \eta_{\min}$. Note that the intervals $[0, \max\{0, h_{\min}(H)\})$ and $[0, \eta_{\min})$ are degenerate if the right endpoint is 0.

For simplicity, we assume that a focal agent's fitness $W(h, H)$ increases as a function of its contribution h if the total good $\eta(h, H)$ is between $\eta_{\min}$ and $\eta_{\max}$,

$$\eta_{\min} < \eta(h, H) < \eta_{\max} \,, \tag{3.8}$$

or, equivalently, if its contribution h lies in the interval

$$\max\{0, h_{\min}(H)\} < h < \max\{0, h_{\max}(H)\} \,. \tag{3.9}$$

Because the fitness benefit $f(\eta(h, H))$ and fitness cost $c(h) = h$ are both increasing functions of h, assuming $W(h, H)$ increases with h means that the benefit of contributing more increases faster than the cost over interval (3.9); equivalently, $f(\eta) - \eta$ is an increasing function of η for $\eta_{\min} < \eta < \eta_{\max}$.

We can now justify our terminology for $\eta_{\min}$ and $\eta_{\max}$. Our assumptions,

A1 fitness is specified by equations (3.3) and (3.5),

A2 $f(\eta)$ is a continuous function defined for $\eta \geq 0$,

A3 $\eta_{\max} > 0$ exists,

A4 if $\eta_{\min}$ exists (which can be assumed without loss of generality) then $0 \leq \eta_{\min} < \eta_{\max}$,

A5 $f(\eta) - \eta$ increases with η when $\eta_{\min} < \eta < \eta_{\max}$, and decreases otherwise,

ensure that for a fixed $H \leq \eta_{\max}/(n-1)$, the focal agent's fitness $W(h, H)$ has a local maximum when the total good $\eta(h, H) = \eta_{\max}$ (*i.e.,* $h = h_{\max}(H) \geq 0$) and a local minimum when the total good $\eta(h, H) = \eta_{\min}$.

Thus, our assumptions describe an n-player snowdrift game with cost $c(h) = h$, and continuous benefit $f(\eta(h, H))$, such that $f(\eta)$ increases faster than linearly in η on a bounded interval, $(\eta_{\min}, \eta_{\max})$, and nowhere else.

3.4 Analysis frameworks

Two frameworks commonly used in analyzing models such as those in the class described in §3.3 are static evolutionary game theory[28, 59, 156] and adaptive dynamics [56, 57, 58, 59]. Below, we recall some of the main concepts from these frameworks, as they apply to our analysis. For a general treatment, see the references cited above.

3.4.1 Static evolutionary game theory

Definition 3.4.1 (Evolutionary stability)**.** *A contribution level $\hat{H} \geq 0$ is (globally)* ***evolutionarily stable*** *(ES)* iff *a single agent that plays a different strategy cannot invade the population (all strategies different from $\hat{H}$ are selected against)[152].*

As different levels of contributions constitute strategies in this game, we also use the term ***evolutionarily stable strategy*** (ESS), when referring to a level of contribution that is ES.

Since evolution by natural selection typically involves mutations that have a small phenotypic effect, the following definition is also biologically relevant:

Definition 3.4.2 (Local Evolutionary stability)**.** *A contribution level $\hat{H} \geq 0$ is* ***locally evolutionarily stable*** *(locally ES) if a single agent playing a mutant strategy h different from, but sufficiently close to $\hat{H}$ cannot invade the population (h is selected against if $|\hat{H} - h|$ is sufficiently small) [59, 157].*

Definition 3.4.3 (Local convergent stability)**.** *A contribution level $\hat{H} \geq 0$ is* ***locally convergently stable*** *(locally CS) if, when the resident strategy H is close enough to $\hat{H}$, a mutant playing a strategy between H and $\hat{H}$ can invade the population (h is selected for if $H < h \leq \hat{H}$ or $\hat{H} \leq h < H$) [158].*

3.4.2 Adaptive dynamics

Adaptive dynamics [56, 57, 58] can also be used to gain insight into similar evolutionary problems. In particular, Doebeli *et al.* [55] use the adaptive dynamics framework to completely characterize the evolutionary dynamics of the continuous snowdrift game with smooth payoffs. Since the class of models defined in § 3.3 is a large subclass of realistic snowdrift games, it is interesting to compare the predictions of [55] to our predictions based on static evolutionary game theory. We therefore briefly outline concepts from adaptive dynamics necessary for this comparison.

Following [55, 57], the ***growth rate*** of a rare mutant strategy h in a resident population

Property	Characterization
Local evolutionary stability	$\frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} < 0$
Convergence stability	$\frac{\partial^2 s_H(h)}{\partial H^2} - \frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} > 0$
Singular strategy can spread in populations playing sufficiently similar strategy	$\frac{\partial^2 s_H(h)}{\partial H^2}\Big\vert_{h=H} > 0$
Mutually-invasible strategies exist near singular point	$\frac{\partial^2 s_H(h)}{\partial H^2} + \frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} > 0$

Table 3.1: Local properties of singular strategies in adaptive dynamics, as in [57, Table 1].

playing H is

$$s_{_H}(h) = W(h,H) - W(H,H)\,, \tag{3.10}$$

where $W(x,y)$ is the fitness of a mutant playing x in a population playing y. The ***local fitness gradient*** is then

$$D(H) = \frac{\partial s_{_H}(h)}{\partial h}\bigg|_{h=H}\,, \tag{3.11}$$

and the ***adaptive dynamics*** of H are given by

$$\dot{H} = D(H)\,. \tag{3.12}$$

An equilibrium of equation (4.32), that is, $\hat{H}$ satisfying $D(\hat{H}) = 0$, is called a ***singular strategy***. A singular strategy that is an attractor of equation (4.32) is convergently stable in the sense of definition 4.A.3. A singular strategy H can also be locally evolutionarily stable as in definition 4.A.2. The mathematical conditions for these and other possible characteristics of singular strategies are listed in table 4.A.1, following [57].

3.5 Results

Below, we summarize our results on the behaviour of the class of models outlined in § 3.3, using static evolutionary game theory and adaptive dynamics. These results are proved in §§3.6, 3.7 and 3.8.

Theorem 3.5.1 (Evolutionary and convergent stability in static theory)**.** *Consider an evolving, infinite, well-mixed population in which fitness is determined by the payoff from a non-linear public goods game played in randomly chosen groups of $n > 1$ agents. Suppose that if the total public good contributed is $\eta \geq 0$, and the benefit to any group member is $f(\eta)$. Thus, if h is a focal agent's contribution and H is the mean non-focal agents'*

contribution to the public good, the focal agent's fitness is

$$W(h,H) = f\big(h+(n-1)H\big) - h\,, \tag{3.13}$$

assuming the cost of the focal agent's contribution is independent of the other member's contributions and contribution is measured in units of its fitness cost. Suppose further that the benefit function f has the following properties:

H1 *f is continuous on $\eta \geq 0$.*

H2 *There exist $\eta_{\min} \geq 0$ and $\eta_{\max} > \eta_{\min}$ such that $f(\eta) - \eta$ increases for $\eta_{\min} < \eta < \eta_{\max}$ and decreases for $\eta < \eta_{\min}$ and $\eta > \eta_{\max}$.*

Then, writing $H^* = \eta_{\max}/n$,

- *If $f(\eta_{\max}) \geq \eta_{\max}$ then the unique ES contribution is H^*.*
- *If $f(\eta_{\max}) < \eta_{\max}$ then*

$$f\big(nH^*\big) - f\big((n-1)H^*\big) \begin{array}{ll} > H^* & \Longrightarrow \text{ two ESSs: } H=0 \text{ and } H=H^*, \\ \leq H^* & \Longrightarrow \text{ unique ESS: } H=0. \end{array} \tag{3.14}$$

Moreover, all existing ESSs are convergently stable.

Remark 3.5.2. *As shown in the proof of theorem 3.5.1, the focal agent's optimal response $h_{\text{opt}}(H)$ (see §3.6.1) is a piecewise-linear function of the mean contribution of the non-focal agents (this can also be seen graphically in figure 3.1).*

Note that if $f(\eta_{\max}) \geq \eta_{\max}$ then it is worthwhile for the focal agent to contribute the maximizing total good, even if it must do so single-handedly (see figure 3.1, first and second panels). The existence of a nonzero ES level of contribution is not surprising in this case. Condition (3.14) says that in the non-trivial situation that $f(\eta_{\max}) < \eta_{\max}$, contributing H^* (*i.e.,* an equal share of the maximizing total good $\eta_{\max}$) is an ESS *iff*, when all nonfocal agents contribute H^*, the cost of contributing H^* (rather than defecting and contributing nothing) is smaller than the increase in the focal agent's benefit resulting from this contribution.[1]

The corresponding analysis based on adaptive dynamics yields:

Theorem 3.5.3 (Local evolutionary and convergent stability in adaptive dynamics)**.** *If the hypotheses of theorem 3.5.1 are satisfied and, in addition,*

H3 *f is twice-differentiable on $\eta \geq 0$,*

[1] Condition (3.14) compares the ***incremental benefit*** of contributing H^* to its cost.

then the adaptive dynamics of H are given by

$$\dot{H} = f'(nH) - 1 \tag{3.15}$$

*and there are two singular points, $H = \eta_{\min}/n$ and $H^* = \eta_{\max}/n$. $H = \eta_{\min}/n$ is a repellor and H^* is an attractor (*i.e., *convergently stable) and a local ESS.*

Figure 3.1 gives the pairwise invasibility plot [57, 58] corresponding to the particular choice of f used in figure 3.1 (see captions for details).

Theorems 3.5.1 and 3.5.3 rely on the assumption of an infinite population to assert that when a mutant arises, the mean fitness of the resident strategy is unaffected by mutant's behaviour [58, 155, 159]. In other words, the average resident does not interact with a mutant. Similarly, it is assumed that the average mutant does not interact with other mutants, so the mean mutant fitness is unaffected by the presence of other mutants (if other mutants exist).

Because real populations are finite, the presence of an invader may well affect the resident's mean fitness, even if the population size is large. For example, in the case of our public goods game, even if there is only one mutant, there are $n-1$ residents in its group, whose fitness is $W(H, \frac{h+(n-2)H}{n-1})$, rather than $W(H, H)$. Thus, in a finite population of size N, even a single invader comprises a nonzero proportion $\epsilon = 1/N > 0$ of the population.

In theorem 3.5.4 below, we relax the assumption that mutants do not affect the resident (or mutant) fitnesses. We retain the assumption of an infinite population, but when considering invasion scenarios, we allow individuals playing the mutant strategy to make up a finite proportion, ϵ, of the population. This new analysis is biologically relevant for at least three reasons:

(i) A mutation might not be selected for when present in a single individual, but spread nevertheless by genetic drift [61]. Once present in a non-negligible proportion of the population, the mutation could be selected for. Thus, we use phenotypic theory to address the question of whether an initially non-adaptive mutation that drifts in to become present in a non-negligible proportion of the population can then be selected for.

(ii) A "bud" consisting of multiple mutants can invade a resident population in a dispersal or migration event (*e.g.,* [160, 161, 162, 163, 164] and references therein), in which case the invading mutants may comprise a non-negligible proportion of the population.

(iii) A mutant may under normal conditions be selected against when rare, but due to an environmental disturbance (either anthropogenic or natural), conditions may temporarily change to allow the mutant to spread (similarly to disturbances of ecological communities, [165, 166]). When the environmental conditions return to normal, the proportion of mutants in the population may have already become non-negligible. In such cases, the invasion analysis must account for more than a single mutant.

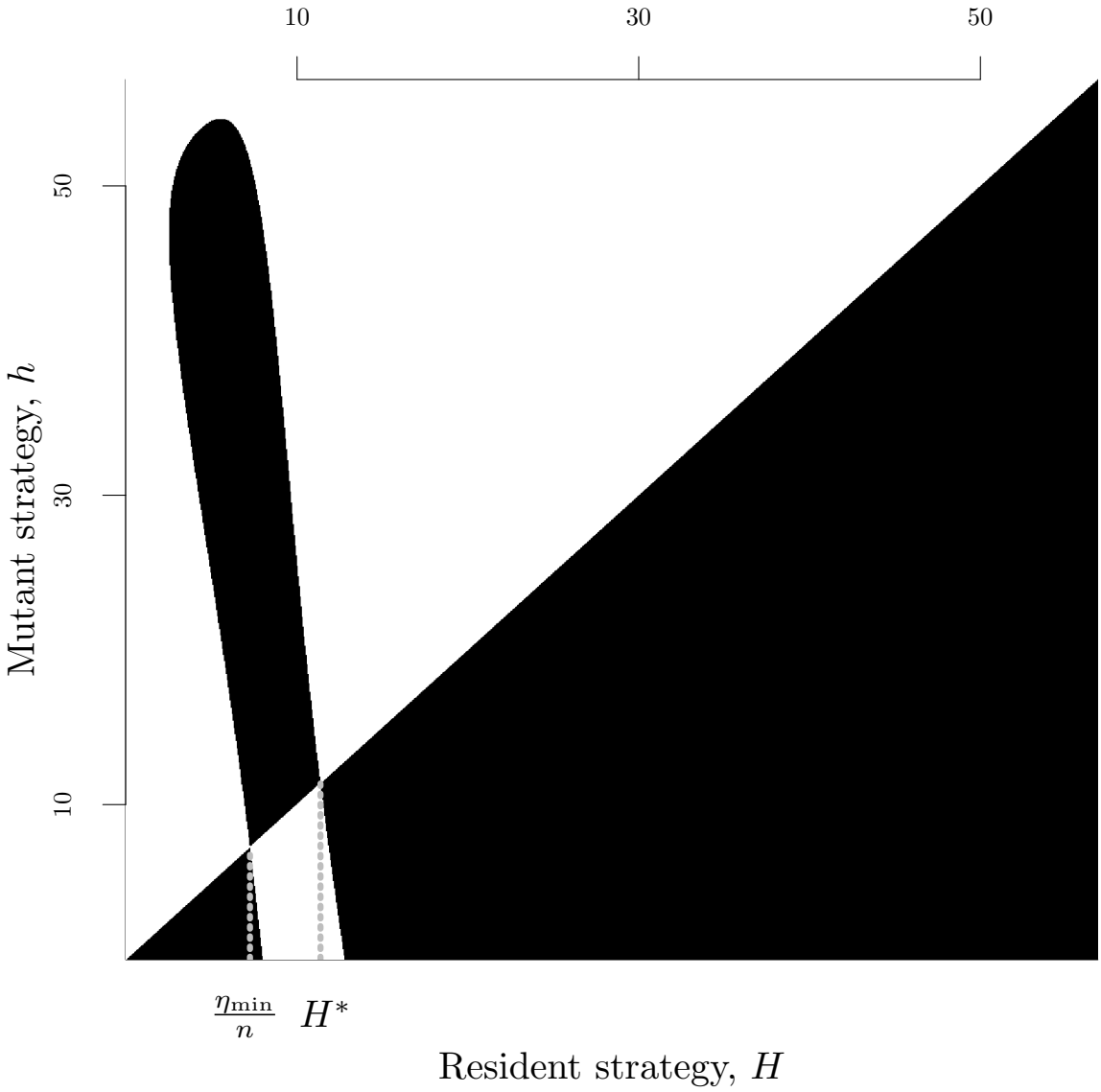

Figure 3.1: Pairwise invasibility plot for sigmoidal benefit $f(\eta)$ with parameters as in figure 3.1: Areas where a single mutant contributing h (vertical axis) can invade an infinite population of agents contributing H (horizontal axis) are shaded. The singular points $H = \eta_{\min}/n$ and $H = H^*$ are marked. The predictions of theorems 3.5.1 and 3.5.3 for this choice of benefit function are that $H = 0$ and $H = H^*$ are the only ESSs and are convergently stable (because $f(\eta_{\max}) < \eta_{\max}$ and $f\big(nH^*\big) - f\big((n-1)H^*\big) > H^*$), and that $\eta_{\min}/n$ is a repellor. The vertical lines at $H = H^*$ and $H = 0$ are unshaded, implying that these are ESS contributions (because no mutant can invade). Near H^*, resident strategies $H \neq H^*$ can be invaded by mutants playing strategies h that are closer to H^*, so H^* is convergently stable, and similarly, so is $H = 0$. $\eta_{\min}/n$ is a repellor, since resident strategies H near $\eta_{\min}/n$ can be invaded by mutants playing h that is farther away from $\eta_{\min}/n$ than H.

For simplicity, we state theorem 3.5.4 with the restriction that $f(\eta_{\max}) < \eta_{\max}$. As noted above, the existence of a nonzero ESS level of contribution when $f(\eta_{\max}) \geq \eta_{\max}$ is trivial. Theorem 3.5.4 then extends the results of theorem 3.5.1 on evolutionary and convergent stability to scenarios where the invading mutants comprise a proportion $\epsilon > 0$ of the population. The result we obtain is weaker, however, in that H^* is no longer guaranteed to be globally ES; it is resistant to invasion by nearby strategies only.

Theorem 3.5.4 (Local evolutionary and convergent stability in static theory with a finite proportion of mutants)**.** *Suppose the hypotheses of theorem 3.5.1 are satisfied and write $H^* = \eta_{\max}/n$. If $f(\eta_{\max}) < \eta_{\max}$ then:*

R1 *(Local ES) If $h \neq H^*$ and a proportion ϵ of mutants playing h arises in a population playing H^* then, if h is sufficiently close to H^*, the mean fitness of a mutant $\overline{W}_{\mathrm{m}}(h)$ is smaller than the mean fitness of a resident $\overline{W}_{\mathrm{r}}(h)$, for any proportion $\epsilon > 0$* (i.e., *the mutants are selected against).*

R2 *(Local CS) If H is sufficiently close to H^*, h is between H and H^*, and a proportion ϵ of mutants playing h arises in a population playing H, then the mean fitness of a mutant $\overline{W}_{\mathrm{m}}(h)$ is greater than the mean fitness of a resident $\overline{W}_{\mathrm{r}}(h)$, for any proportion $\epsilon > 0$* (i.e., *the mutants are selected for).*

3.6 Proof of theorem 3.5.1

Without loss of generality, we can assume that

$$f(0) = 0\,. \tag{3.16}$$

To see this, note that the dynamics of the game are not changed by adding a constant to the fitness function $W(h, H)$. If $f(0) \neq 0$ then we would analyze $\tilde{f}(\eta) = f(\eta) - f(0)$ (so $\tilde{f}(0) = 0$) and $\tilde{W}(h, H) = W(h, H) - f(0) = \tilde{f}(\eta(h, H)) - h$.

The structure of our proof of theorem 3.5.1 is as follows: In §3.6.1, we find the optimal response for the focal agent as a function of the non-focal agents' mean contribution, H. Then, in §3.6.2, we use the focal agent's optimal response to show that the only possible ESSs are either not contributing ($H = 0$), or contributing an equal share of the maximizing total good, $H^* = \eta_{\max}/n > 0$. Lastly, in §3.6.3, we show that these ESSs are also convergently stable.

3.6.1 Optimal response for focal agent

For a given mean contribution from the non-focal agents, H, what must the focal agent contribute in order to maximize its fitness? This is the focal agent's ***optimal response*** to H,

which we denote $h_{\text{opt}}(H)$. The optimal response in the class of games we are considering is given in lemma 3.6.1 and plotted in figure 3.1.

Lemma 3.6.1 (Best response lemma)**.** *Under the conditions of theorem 3.5.1*

- *if* $f(\eta_{\max}) > \eta_{\max}$*, then*

$$h_{\text{opt}}(H) = \max\left\{0, h_{\max}(H)\right\} = \begin{cases} h_{\max}(H) & 0 \leq H < \frac{\eta_{\max}}{n-1}\,, \\ 0 & \frac{\eta_{\max}}{n-1} \leq H\,, \end{cases} \tag{3.17a}$$

  where $h_{\max}(H)$ *is defined in equation* (3.6b)*. Note that the interval* $[0, \eta_{\max}/(n-1))$ *is never empty, because* $\eta_{\max} > 0$*.*

- *if* $f(\eta_{\max}) \leq \eta_{\max}$*, then,*

$$h_{\text{opt}}(H) = \begin{cases} 0 & & 0 \leq H < H^0\,, \\ 0 \quad \text{or} \quad h_{\max}(H^0) & & H = H^0 \\ h_{\max}(H) & & H^0 < H < \frac{\eta_{\max}}{n-1}\,, \\ 0 & & \frac{\eta_{\max}}{n-1} \leq H\,, \end{cases} \tag{3.17b}$$

  where H^0 *is the unique solution of*

$$f\big((n-1)H\big) - (n-1)H = f(\eta_{\max}) - \eta_{\max} \tag{3.18}$$

  such that $0 \leq H^0 < \frac{\eta_{\min}}{n-1}$*.*

In equation (3.17b)*, note that the first interval* $(0 \leq H < H^0)$ *is empty if* $H^0 = 0$*, and that for* $H = H^0$*,* $h_{\text{opt}}(H)$ *is 2-valued.*

Proof. Observe that depending on the mean non-focal agents' contribution H, $W(h, H)$ behaves in one of the following ways:

1. If the non-focal agents' contribution is no less than the maximizing total good ($(n-1)H \geq \eta_{\max}$), then the focal agent's fitness $W(h, H)$ decreases with its contribution, h. The optimal contribution for the focal agent is then $h_{\text{opt}}(H) = 0$.

2. If the non-focal agents' contribution is greater than or equal to the minimizing total good and smaller than the maximizing total good ($\eta_{\min} \leq (n-1)H < \eta_{\max}$), then the focal agent's fitness $W(h, H)$ increases for $0 \leq h \leq h_{\max}(H)$ and decreases for $h > h_{\max}(H)$, so the focal agent's optimal contribution is $h_{\text{opt}}(H) = \eta_{\max} - (n-1)H$.

3. If the non-focal agents' contribution is lower than the minimizing total good ($(n-1)H < \eta_{\min}$), fitness decreases for $0 \leq h \leq h_{\min}(H)$, increases for $\max\{0, h_{\min}(H)\} < h \leq h_{\max}(H)$ and decreases again for $h > h_{\max}(H)$.

   Thus, two levels of contribution locally maximize the focal agent's fitness: $h =$

$\eta_{\max} - (n-1)H$ and $h = 0$, for which the focal agent's fitness is $W(\eta_{\max} - (n-1)H, H) = f(\eta_{\max}) - \eta_{\max} + (n-1)H$ and $W(0, H) = f((n-1)H)$, respectively. These two local fitness maxima are the candidates for the global fitness maximum, that is, the focal agent's optimal response $h_{\text{opt}}(H)$. Let

$$\begin{aligned} \Delta W(H) &= W(\eta_{\max} - (n-1)H, H) - W(0, H) \\ &= f(\eta_{\max}) - \eta_{\max} - [f((n-1)H) - (n-1)H] \end{aligned} \tag{3.19}$$

be the difference between the focal agent's two local fitness maxima. Note that since $f(\eta) - \eta$ decreases with η on $[0, \eta_{\min}]$, $\Delta W(H)$ is an increasing function of H on $\left[0, \frac{\eta_{\min}}{n-1}\right)$.

Because $f(\eta) - \eta$ increases with η on $[\eta_{\min}, \eta_{\max}]$, it follows that

$$\Delta W\left(\frac{\eta_{\min}}{n-1}\right) = f(\eta_{\max}) - \eta_{\max} - [f(\eta_{\min}) - \eta_{\min}] > 0\,. \tag{3.20}$$

Thus, since $\Delta W(H)$ is continuous, it follows that for large enough values of $H < \eta_{\min}/(n-1)$, $\Delta W(H) > 0$, implying that the focal agent maximizes fitness by contributing $h_{\text{opt}}(H) = h_{\max}(H) > 0$.

At the other extreme end of the interval $0 \leq H < \eta_{\min}/(n-1)$, we have:

$$\Delta W(0) = f(\eta_{\max}) - \eta_{\max}\,. \tag{3.21}$$

There are three possible cases:

(i) If $\Delta W(0) > 0$, then for all $0 \leq H < \eta_{\min}/(n-1)$, the focal agent's optimal response is $h_{\text{opt}}(H) = h_{\max}(H) > 0$.

The condition $\Delta W(0) > 0$ has a simple biological interpretation: If $\Delta W(0) > 0$, then the benefit to the focal agent when the total good is equal to the total maximizing good ($\eta = \eta_{\max}$) is so large that—even if the non-focal group members contribute nothing—the focal agent gains by single-handedly contributing the total maximizing good ($h = \eta_{\max}$). It is then sensible that if the nonfocal agents have collectively contributed less than the minimizing total good $(n-1)H < \eta_{\min}$ (or indeed, less than the maximizing total good, $\eta_{\max}$), then the focal agent still does best by ensuring that the total maximizing good is attained ($\eta = \eta_{\max}$) (in fact, if the non-focal agents contribute $0 < (n-1)H < \eta_{\max}$, the focal agent's fitness must be higher than when $H = 0$, since it is now required to contribute less to obtain the same benefit).

(ii) If $\Delta W(0) < 0$, then because $\Delta W(H)$ is continuous, increasing, and $\Delta W\left(\frac{\eta_{\min}}{n-1}\right) >$

0, it follows that there is a unique solution to

$$\Delta W(H^0) = 0, \qquad 0 < H^0 < \frac{\eta_{\min}}{n-1}\,. \tag{3.22}$$

Moreover, $\Delta W(H) < 0$ for $H < H^0$ and $\Delta W(H) > 0$ for $H > H^0$.

Thus, the optimal response for the focal agent is $h_{\text{opt}}(H) = 0$ if $H < H^0$, and $h_{\text{opt}}(H) = h_{\max}(H) > 0$ if $H^0 < H < \frac{\eta_{\min}}{n-1}$. If $H = H^0$, the focal agent can maximize its fitness by contributing either $h = 0$ or $h = \eta_{\max} - (n-1)H^0 > 0$ (because $W(0, H^0) = W(\eta_{\max} - (n-1)H^0, H^0)$).

H^0 is the mean non-focal agents' contribution for which the focal agent obtains the same fitness either by contributing nothing ($h = 0$), or by completing the difference between the maximizing total contribution and the collective contribution of the non-focal agents (so that $\eta = \eta_{\max}$).

(iii) If $\Delta W(0) = 0$, then $\Delta W(H) > 0$ for all $H > 0$, and $h_{\text{opt}}(H) = h_{\max}(H) > 0$ for $0 < H < \eta_{\min}$. For $H = 0$, the focal agent can maximize its fitness by contributing either $h = 0$ or $h = \eta_{\max} > 0$ (again, $W(0,0) = W(\eta_{\max}, 0)$).

□

3.6.2 Evolutionarily stable contribution levels

To interpret definition 5.6.1 for evolutionary stability mathematically, suppose that the entire population consists of agents playing H. The focal agent's fitness is given by equation (3.5) where h can be an alternative strategy $h \neq H$. The fitness of the $n-1$ non-focal individuals in the focal agent's group is

$$W\left(H, \frac{h + (n-2)H}{n-1}\right) = f\big(\eta(h, H)\big) - H\,. \tag{3.23}$$

Thus, the focal agent's fitness is larger than that of the non-focal individuals in its group *iff* $h < H$. However, since the population is infinitely large, an infinite number of non-focal individuals are in homogeneous groups in which everyone contributes H, so their fitness is $W(H, H)$. Thus, the mean fitness of a non-focal individual remains $W(H, H)$. Then, $H = H^*$ is an ESS *iff* when all non-focal agents play H^*, if the focal agent plays an alternative strategy $h \neq H^*$, its fitness is lower than the resident strategy H^*, or: $W(h, H^*) < W(H^*, H^*)$ for all $h \neq H^*$.

Thus, $H^* \geq 0$ is an ESS if and only if it is the unique optimal response to itself. Explicitly, H^* must solve

$$h_{\text{opt}}(H^*) = H^*\,, \tag{3.24}$$

and $h_{\text{opt}}(H^*)$ must be univalued. H for which $h_{\text{opt}}(H)$ is not univalued cannot be an ESS even if one of the values of $h_{\text{opt}}(H)$ is H, because when the non-focal agents play H, the focal agent can play an alternative strategy (one of the other values of $h_{\text{opt}}(H)$ without decreasing its fitness). Geometrically, solutions of equation (3.24) are intersections (in the H–h plane) of the curve $h = h_{\text{opt}}(H)$ with the line $h = H$.

Note that while solutions of equation (3.24) at which h_{opt} is not univalued are not ESSs, they are still technically Nash Equilibria [86].

We separate the discussion into the following cases:

1. $f(\eta_{\max}) > \eta_{\max}$:

   In this case, $h_{\text{opt}}(H)$ is given by equation (3.17a) (see figure 3.1, top panel). Solving equation (3.24) yields a unique ESS,

   $$H^* = \frac{\eta_{\max}}{n}\,. \tag{3.25}$$

   Note that when $f(\eta_{\max}) > \eta_{\max}$, it is beneficial for the focal agent to ensure the maximizing total good is attained even if it must do so single-handedly (so there is no "tragedy of the commons" in this scenario). Thus, it is biologically sensible that at the ESS all group members contribute equally towards the maximizing total good.

2. $f(\eta_{\max}) = \eta_{\max}$:

   In this case, $h_{\text{opt}}(H)$ is given by equation (3.17b) with $H^0 = 0$, that is,

   $$h_{\text{opt}}(H) = \begin{cases} 0 \quad \text{or} \quad \eta_{\max} & H = 0\,, \\ h_{\max}(H) & 0 < H < \frac{\eta_{\max}}{n-1}\,, \\ 0 & H \geq \frac{\eta_{\max}}{n-1}\,. \end{cases} \tag{3.26}$$

   Thus, $H = 0$ is not an ESS, because if the non-focal agents contribute $H = 0$, the focal agent's fitness if it contributes $h = \eta_{\max} > 0$ is identical to its fitness if it does not contribute ($h = 0$).

   If $H \geq \frac{\eta_{\max}}{n-1}$, then $h_{\text{opt}}(H) = 0 < H$ and so H cannot be an ESS. Lastly, if $0 < H < \frac{\eta_{\max}}{n-1}$ then solving equation (3.24) again yields a unique ESS given by equation (3.25) (see Figure 3.1, second panel).

   The biological interpretation of the ESS $H^* = \eta_{\max}/n$ is similar to the previous case ($f(\eta_{\max}) > \eta_{\max}$). The only change is that now, $H = 0$ (no contribution) is a NE (where it was not when $f(\eta_{\max}) > \eta_{\max}$), because when the non-focal agents do not contribute, the focal agent's fitness can be maximized either by contributing $\eta_{\max}$ or not contributing. By contrast, $H^* = \eta_{\max}/n$ is still ES, because when all group members contribute H^*, then the focal agent's fitness will decrease if it contributes

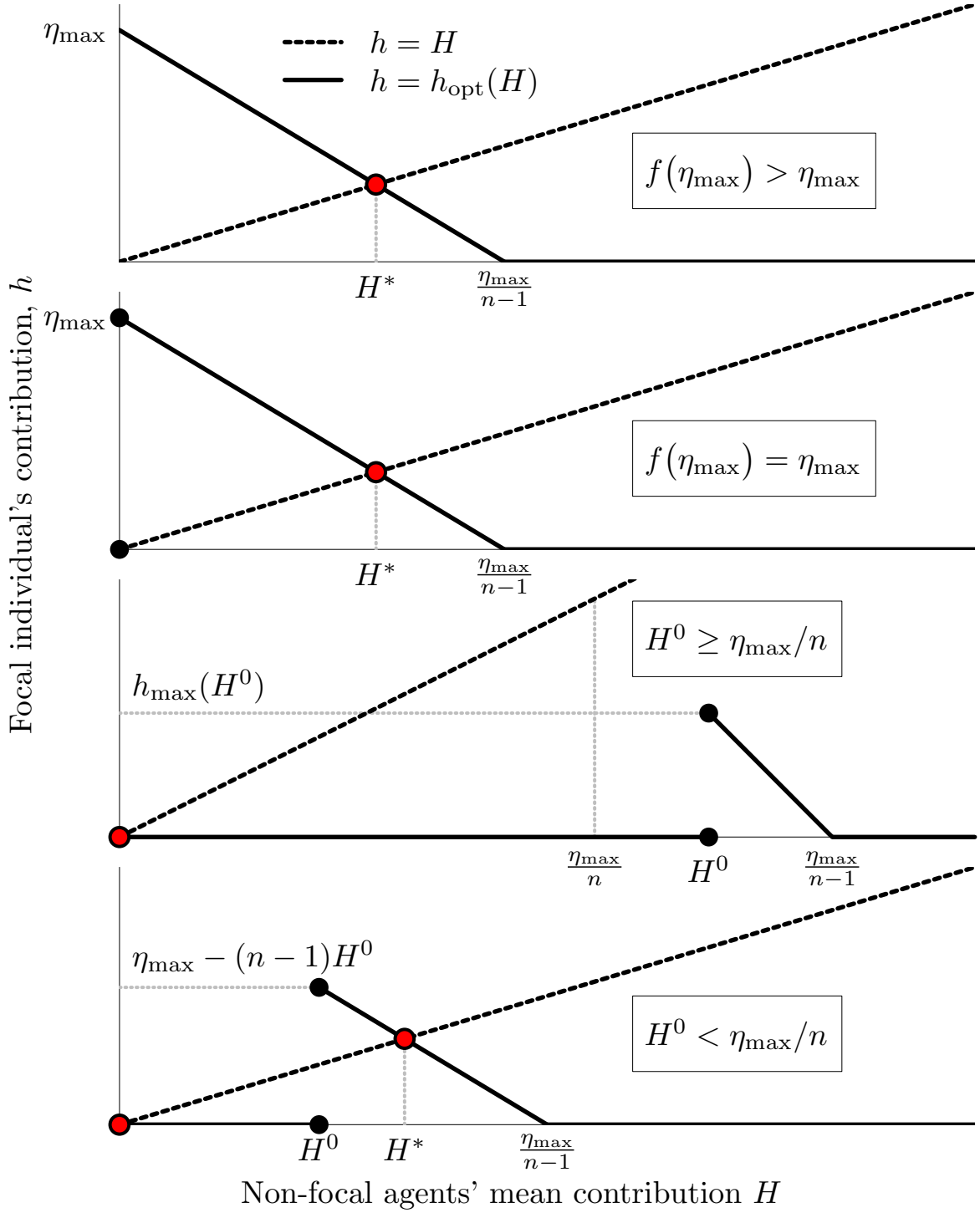

Figure 3.1: ESSs (red dots) are solutions of $H = h_{\text{opt}}(H)$ [equation (3.24)] at which $h_{\text{opt}}(H)$ is univalued. $H = h_{\text{opt}}(H)$ implies $h = H$ is a best responses to itself, and $h_{\text{opt}}(H)$ being univalued ensures that no other invading strategy matches the resident's fitness. The four panels (top to bottom), depict the intersections of $h = H$ (dashed black line) and $h = h_{\text{opt}}(H)$ (solid black line) in the h-H plane in the four possible cases: $f(\eta_{\max}) > \eta_{\max}$, $f(\eta_{\max}) = \eta_{\max}$, $H^0 \geq \eta_{\max}/n$, and $H^0 < \eta_{\max}/n$. For values of H (the mean nonfocal agents' help) where the focal agent's best response $h_{\text{opt}}(H)$ is two-valued, its values are indicated by black dots.

$h \neq H^*$.

3. $\underline{f(\eta_{\max}) < \eta_{\max}}$: In this case, h_{opt} is given by equation (3.17b), so $H = 0$ is ES.

   For $0 < H < H^0$ and $H \geq \frac{\eta_{\max}}{n-1}$, $h_{\text{opt}}(H) = 0 < H$, so Equation (3.24) is not satisfied, and H cannot be ES. Likewise, $H = H^0$ cannot be an ESS, since $h_{\text{opt}}(H^0)$ is not univalued.

   However, depending on the relationship between $\eta_{\max}$, n and H^0, there may or may not be another ESS in the range $H^0 < H < \frac{\eta_{\max}}{n-1}$:

   (a) $\underline{H^0 \geq \eta_{\max}/n}$: When $H^0 \geq \eta_{\max}/n$, there is no additional (nonzero) ESS in the range $H^0 < H < \frac{\eta_{\max}}{n-1}$, because

   $$h_{\text{opt}}(H) < \eta_{\max} - (n-1)H^0 < \frac{\eta_{\max}}{n} < H\,, \tag{3.27}$$

   (recall that $h_{\text{opt}}(H)$ decreases linearly with H in this range; see Figure 3.1, third panel). Thus, for this sub-case, $H^* = 0$ is the unique ESS.

   The condition $H^0 \geq \eta_{\max}/n$ is equivalent to $\Delta W(\eta_{\max}/n) \leq 0$, or

   $$\begin{aligned} W\left(\frac{\eta_{\max}}{n}, \frac{\eta_{\max}}{n}\right) &= f\left(\eta_{\max}\right) - \frac{\eta_{\max}}{n} \\ &\leq f\left(\frac{n-1}{n}\eta_{\max}\right) = W\left(0, \frac{\eta_{\max}}{n}\right)\,. \end{aligned} \tag{3.28}$$

   Condition (3.28) simply states that if all agents contribute equally towards achieving the maximizing total good, the focal agent does no worse by withdrawing its contribution (*i.e.,* switching to $h = 0$).

   (b) $\underline{H^0 < \eta_{\max}/n}$:

   To see that in this case, there is a second (nonzero) ES level of contribution H, we seek a solution of Equation (3.24) in the range $H^0 < H < \frac{\eta_{\max}}{n-1}$. In this range, $h_{\text{opt}}(H) = \eta_{\max} - (n-1)H$, so $H^* = \eta_{\max}/n$ solves equation (3.24) ($H^0 < H^* = \eta_{\max}/n$ by our assumption for this sub-case, and $H^* = \eta_{\max}/n < \eta_{\max}/(n-1)$ trivially, so $H^* \in \left(H^0, \frac{\eta_{\max}}{n-1}\right)$ as required).

   Thus, there are in this case two ESS contribution levels: $H = 0$ and $H = \frac{\eta_{\max}}{n}$ (see figure 3.1, bottom panel).

   To understand why there is an additional (non-zero) ES level of contribution when $H^0 < \eta_{\max}/n$, note that by definition, $H^0 < \frac{\eta_{\min}}{n-1}$ (see equation (3.22)), so there are two possibilities:

   (i) If $\eta_{\max}/n < \eta_{\min}/(n-1)$ then $H^0 < \eta_{\max}/n$ *iff* $\Delta W\left(\frac{\eta_{\max}}{n}\right) > 0$, or

equivalently,

$$f(\eta_{\max}) - f(\frac{n-1}{n}\eta_{\max}) > \frac{\eta_{\max}}{n}\,, \tag{3.29}$$

which is the converse of condition (3.28).

The biological intuition for this case is that if $\eta_{\max}/n < \eta_{\min}/(n-1)$, then when all non-focal agents contribute $H^* = \eta_{\max}/n$, their total contribution is less than the minimizing total good ($(n-1)H^* < \eta_{\min}$). Consequently, $W(h, H^*)$ decreases for $0 \leq h \leq \eta_{\min} - (n-1)H^*$, then increases for $\eta_{\min} - (n-1)H^* < h < \eta_{\max} - (n-1)H^*$ and decreases again for $h \geq \eta_{\max} - (n-1)H^*$. Thus, the two candidates for the best response for the focal agent are $h = 0$ and $h = \eta_{\max} - (n-1)H^* = H^*$, and H^* is an ESS *iff* $W(H^*, H^*) > W(0, H^*)$ (that is, $\Delta W(H^*) > 0$).

Note that condition (3.29) stipulates that the mean slope of f on the interval $\left[\frac{n-1}{n}\eta_{\max}, \eta_{\max}\right]$ is larger than 1.

(ii) If $\eta_{\min}/(n-1) \leq \eta_{\max}/n$, then $H^0 < \eta_{\max}/n$ is satisfied (because $H^0 < \eta_{\min}/n$). This is because if $\eta_{\min}/(n-1) \leq \eta_{\max}/n$, then when $H = H^* = \eta_{\max}/n$, the total non-focal agents' contribution exceeds the minimizing total good ($(n-1)H > \eta_{\min}$), so $W(h, H^*)$ is unimodal and has a unique global maximum (*i.e.,* in the range $h \geq 0$) at

$$h = \eta_{\max} - (n-1)H^* = \frac{\eta_{\max}}{n} = H^*\,. \tag{3.30}$$

While the condition $\eta_{\min}/(n-1) \leq \eta_{\max}/n$ seems weaker than condition (3.29), note that $W(0, H^*) < W(H^*, H^*)$, so condition (3.29) must hold in this case as well.

We conclude that if $f(\eta_{\max}) < \eta_{\max}$, then $H = 0$ is an ESS, and additionally, $H^* = \eta_{\max}/n$ is an ESS *iff* condition (3.29) holds.

Also, note that for a fixed benefit function, $f(\eta)$, for sufficiently large group size n, $\eta_{\min}/(n-1) \leq \eta_{\max}/n$ *must* hold. Thus, all else being equal, larger groups are more likely to have a nonzero ESS contribution.

3.6.3 Convergent stability of the ESSs

§3.6.2 showed that unless both $f(\eta_{\max}) < \eta_{\max}$ and $f(\eta_{\max}) \leq f(\frac{n-1}{n}\eta_{\max}) + \eta_{\max}/n$, the contribution level $H^* = \eta_{\max}/n$ is ES. We now show that when $H^* = \eta_{\max}/n$ is an ESS, it is also convergently stable, that is:

Suppose that all members of the population contribute H, and that a mutant playing

$h \neq H$ enters the population. Recalling that we assume an infinite population, the mean resident fitness is unaffected by the mutant, and is simply $W(H,H)$. Thus, we wish to show that if H is sufficiently close to H^*, then for any invading strategy h that is between H and H^*, $W(h,H) > W(H,H)$.

Suppose that the resident strategy is $H = H^* - \mu$, where $\mu > 0$, and that the mutant strategy satisfies $H^* - \mu = H < h < H^*$. If $\mu < (\eta_{\max} - \eta_{\min})/n$ then $\eta_{\min} < nH < h + (n-1)H\eta_{\max}$, so

$$\begin{aligned} W(h,H) - W(H,H) &= [f(h + (n-1)H) - h] - [f(nH)) - H] \\ &= [f(h + (n-1)H) - (h + (n-1)H)] \\ &\quad - [f(nH)) - nH] > 0 \end{aligned} \tag{3.31}$$

because $f(\eta) - \eta$ is increasing on $[\eta_{\min}, \eta_{\max}]$.

Now suppose that the resident strategy is $H = H^* + \mu$, where $\mu > 0$. Because $\eta_{\max} < h + (n-1)H < nH$ and $f(\eta) - \eta$ decreases for $\eta > \eta_{\max}$, we have

$$\begin{aligned} W(h,H) - W(H,H) &= [f(h + (n-1)H) - (h + (n-1)H)] \\ &\quad - [f(nH)) - nH] > 0\,, \end{aligned} \tag{3.32}$$

for *any* $\mu > 0$.

It follows that if all members of the group use a strategy H sufficiently near the non-zero equilibrium, H^*, then the fitness of a mutant strategy between H and H^* is larger than the mean resident fitness, so H^* is convergently stable.

We also saw in §3.6.2, that if $f(\eta_{\max}) < \eta_{\max}$ then $H = 0$ is ES. To see that it is also convergently stable, note that if $0 < H < \eta_{\min}/n$, then $\eta(h,H) \leq \eta(H,H) < \eta_{\min}$ for all for $0 < h \leq H$. Since $f(\eta) - \eta$ decreases with η for $\eta < \eta_{\min}$, and $\eta(h,H)$ increases with h, it follows that $W(h,H) = f(\eta(h,H)) - \eta(h,H) + (n-1)H$ decreases with h, which implies $H = 0$ is convergently stable.

3.7 Proof of theorem 3.5.3

Following [55, 57], the growth rate of a rare mutant strategy h in a resident population playing H is

$$s_{_H}(h) = W(h,H) - W(H,H) = f(h + (n-1)H) - f(nH) + H - h\,. \tag{3.33}$$

The local fitness gradient is then

$$D(H) = \left.\frac{\partial s_{_H}(h)}{\partial h}\right|_{h=H} = f'(nH) - 1\,, \tag{3.34}$$

and the adaptive dynamics of H are given by

$$\dot{H} = D(H) = f'(nH) - 1\,. \tag{3.35}$$

Singular strategies satisfy $f'(nH) = 1$. Since by our assumptions, $f(\eta) - \eta$ increases when $\eta_{\min} < \eta < \eta_{\max}$ and decreases otherwise, the two singular strategies are $H = \eta_{\min}/n$ and $H = \eta_{\max}/n = H^*$. Because $\frac{d}{dH}D(H) = nf''(nH)$, $f''(\eta_{\min}) > 0$, it follows that $H = \eta_{\min}/n$ is a repellor.

As for the singular strategy $H = H^*$, using table 4.A.1 (adapted from [57]) and letting

$$\begin{aligned} a &= \left.\frac{\partial^2 s_{_H}(h)}{\partial^2 H}\right|_{h=H=H^*} = (n-1)^2 f''(nH^*) - n^2 f''(nH^*) \\ &= (1-2n) f''(\eta_{\max}) > 0\,, \end{aligned} \tag{3.36}$$

$$b = \left.\frac{\partial^2 s_{_H}(h)}{\partial^2 h}\right|_{h=H=H^*} = f''(nH^*) = f''(\eta_{\max}) < 0\,, \tag{3.37}$$

(since $f''(\eta_{\max}) < 0$ and $n \geq 1$) we see that H^* is a convergently stable local-ESS (because $b < 0$ and $a > b$). Though these are the only properties necessary for theorem 3.5.3, for completeness, we also note that since $a > 0$, H^* can invade a homogeneous population playing a nearby strategy $H \neq H^*$. Lastly, if $n > 1$, then mutually-invasible strategies exist near H^* since $a + b > 0$ (however, dimorphic populations will tend to disappear as the population converges towards the ESS H^*, see [57, p.42]).

3.8 Proof of theorem 3.5.4

3.8.1 Local ES

Proof of R1 (Local ES). Consider an infinite population playing H^* invaded by a proportion $\epsilon > 0$ of mutants playing $h \neq H^*$. We wish to compare the mean fitnesses of a resident playing H^* and a mutant playing h.

To obtain the mean mutant fitness, first note that the payoff to a mutant in a group with

a total of k mutants is

$$W_{\mathrm{m},k}(h) = W\left(h, \frac{(k-1)h + (n-k)H^*}{n-1}\right). \tag{3.38}$$

We now calculate the proportion of mutants that are in a group containing k mutants. Choose an agent at random from the population by first choosing a group at random and then choosing an agent at random from within that group. Let I be an indicator for whether the chosen agent is a mutant ($I = 1$ if the chosen agent is a mutant, and $I = 0$ otherwise). Let M be the number of mutants in the chosen group. We use Bayes' Theorem [167] to find $P(M = k|I = 1)$, that is, the probability that a chosen mutant is in a group containing k mutants:

$$P(M = k|I = 1) = \frac{P(M = k)P(I = 1|M = k)}{P(I = 1)}. \tag{3.39}$$

Because the population is assumed infinite, the probability that a randomly chosen group contains k mutants is binomially distributed with parameters n and ϵ,

$$P(M = k) = \binom{n}{k}\epsilon^k(1-\epsilon)^{n-k}. \tag{3.40}$$

The probability of drawing a mutant at random from a group containing k mutants is $P(I = 1|M = k) = k/n$. The probability that an individual chosen at random from the whole population is a mutant is $P(I = 1) = \epsilon$. Thus,

$$\begin{aligned} P(M = k|I = 1) &= \frac{\binom{n}{k}\epsilon^k(1-\epsilon)^{n-k}k/n}{\epsilon} \\ &= \binom{n-1}{k-1}\epsilon^{k-1}(1-\epsilon)^{n-k}. \end{aligned} \tag{3.41}$$

that is, the remaining number of mutants in the group is distributed binomially with parameters $n-1$ and ϵ.

It follows that the mean payoff for a mutant is:

$$\begin{aligned} \overline{W}_{\mathrm{m}}(h) &= \sum_{k=1}^{n} P(M = k|I = 1)W_{\mathrm{m},k}(h) \\ &= \sum_{k=1}^{n} \binom{n-1}{k-1}\epsilon^{k-1}(1-\epsilon)^{n-k}W_{\mathrm{m},k}(h) \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k}\epsilon^{k}(1-\epsilon)^{n-1-k}W_{\mathrm{m},k+1}(h) \end{aligned}$$

Similarly, the probability that a randomly chosen resident's group contains k mutants is

$$P(M = k|I = 0) = \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k}\,, \tag{3.42}$$

and the payoff to a resident in a group containing k mutants is (equation (3.5))

$$W_{\mathrm{r},k} = W\left(H^*, \frac{kh + (n-1-k)H^*}{n-1}\right) = f\left(kh + (n-k)H^*\right) - H^*\,. \tag{3.43}$$

So, the mean payoff to an agent playing the resident strategy H^* is

$$\begin{aligned}\overline{W}_{\mathrm{r}}(h) &= \sum_{k=0}^{n-1} P(M = k|I = 0) W_{\mathrm{r},k}(h) \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k} W_{\mathrm{r},k}(h)\,.\end{aligned}$$

The difference between the mean fitnesses of the mutant and resident strategies is then

$$\begin{aligned}\delta\overline{W}(h) &= \overline{W}_{\mathrm{m}}(h) - \overline{W}_{\mathrm{r}}(h) \\ &= \sum_{k=0}^{n-1} \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k} \Big[W_{\mathrm{m},k+1}(h) - W_{\mathrm{r},k}(h)\Big]\,.\end{aligned} \tag{3.44}$$

Denoting the total contribution in a group containing k mutants and $n-k$ residents by

$$\eta_k = kh + (n-k)H^*\,, \tag{3.45}$$

and noting that $\eta_{k+1} - \eta_k = h - H^*$, we have

$$\begin{aligned}W_{\mathrm{m},k+1}(h) - W_{\mathrm{r},k}(h) &= \big[f(\eta_{k+1}) - h\big] - \big[f(\eta_k) - H^*\big] \\ &= \big[f(\eta_{k+1}) - \eta_{k+1}\big] - \big[f(\eta_k) - \eta_{k+1}\big]\,.\end{aligned} \tag{3.46}$$

If $\frac{\eta_{\min}}{n} < h < H^*$ then, for all $0 \le k \le n-1$,

$$\eta_{\min} < nh = \eta_n \le \eta_{k+1} < \eta_k \le \eta_0 = nH^* = \eta_{\max}\,, \tag{3.47}$$

so because $f(\eta) - \eta$ is increasing for $\eta_{\min} < \eta < \eta_{\max}$,

$$W_{\mathrm{m},k+1}(h) - W_{\mathrm{r},k}(h) = \big[f(\eta_{k+1}) - \eta_{k+1}\big] - \big[f(\eta_k) - \eta_{k+1}\big] < 0\,, \tag{3.48}$$

that is, each term in the sum (3.44) above is negative, implying $\delta\overline{W}(h) < 0$.

Similarly, if $h > H^*$ then, for all $0 \leq k \leq n-1$,

$$\eta_{\max} = nH^* < \eta_k < \eta_{k+1}\,, \tag{3.49}$$

and since $f(\eta) - \eta$ is decreasing for $\eta > \eta_{\max}$, inequality (3.48) holds again, implying $\delta\overline{W}(h) < 0$.

Thus, a mutant strategy sufficiently close to the equilibrium H^* cannot invade, regardless of its initial proportion in the population, ϵ. □

3.8.2 Local CS

Proof of R2 (Local CS). Similar to the derivation of equation (3.44) in § 3.8.1, the mean fitness difference between a mutant contributing h and a resident contributing H is

$$\begin{aligned}
\delta\overline{W}(h) &= \sum_{k=0}^{n-1} \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k} \left[W\left(h, \frac{kh + (n-1-k)H}{n-1}\right) \right. \\
&\qquad \left. - W\left(H, \frac{kh + (n-1-k)H}{n-1}\right) \right] \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k} \left\{ \left[f\big((k+1)h + (n-1-k)H\big) - h \right] \right. \\
&\qquad \left. - \left[f\big(kh + (n-k)H\big) - H \right] \right\} \\
&= \sum_{k=0}^{n-1} \binom{n-1}{k} \epsilon^k (1-\epsilon)^{n-1-k} \times \\
&\qquad \left\{ \left[f\big((k+1)h + (n-1-k)H\big) - \big((k+1)h + (n-1-k)H\big) \right] \right. \\
&\qquad \left. - \left[f\big(kh + (n-k)H\big) - \big(kh + (n-k)H\big) \right] \right\}.
\end{aligned} \tag{3.50}$$

If $\eta_{\min}/n < H < h < H^*$, then $\eta_{\min} < kh + (n-k)H < \eta_{\max}$ for all $0 \leq k \leq n$, and because $f(\eta) - \eta$ is increasing on $[\eta_{\min}, \eta_{\max}]$, each term in the sum in the last line of equation (3.50) is positive, so $\delta\overline{W}(h) > 0$. Conversely, if $H^* < h < H$, then because $\eta_{\max} < kh + (n-k)H$ for all $0 \leq k \leq n$ and $f(\eta) - \eta$ decreases for $\eta > \eta_{\max}$, again, $\delta\overline{W}(h) > 0$. Thus, the ESS H^* is convergently stable. □

3.9 Discussion

We have analyzed the general class of public goods games described in § 3.3 (continuous n-player snowdrift games [28, 55]) using the two frameworks summarized in § 3.4 (static evolutionary game theory [28, 59, 156] and adaptive dynamics [56, 57, 58, 59]). Our results are expressed in three theorems stated in § 4.3 and proved in §§3.6, 3.7 and 3.8.

With the standard static theory, we identified two candidate evolutionarily stable strategies (ESSs): either contributing nothing ($H = 0$, "defection") or contributing an equal share of the maximizing total good ($H = H^* = \eta_{\max}/n$, "cooperation"). Defection is an ESS unless the benefit of contributing to the public good is so large that it is worth doing so even if no-one else contributes. Cooperation is an ESS unless the cost of contributing the maximizing total good *single-handedly* outweighs its benefit and the incremental cost of contributing an equal share also exceeds its benefit (condition (3.14)). When they exist, each ESS is resistant to invasion by a mutant that contributes *any* other amount (globally evolutionary stable) and is selected for in populations of individuals contributing nearly the ESS level (locally convergently stable).

Our conclusions do not depend on the form of the benefit function $f(h)$ beyond the biologically sensible hypotheses H1 and H2, and are hence applicable to a wide range of biological, social and economic problems. Moreover, the conditions we find under which cooperation is inherently evolutionarily stable are independent of any external mechanism such as population structure [117, 160, 168], kin selection [138, 160, 163, 169], reciprocity [139, 170, 171] or partner selection [172].

With the adaptive dynamics framework, we found only one possible ESS, which is to "cooperate" by contributing an equal share of the maximizing total good, as in the static theory. Unlike the static theory, the standard formulation of adaptive dynamics requires a smooth fitness function (hypothesis H3) and the typical definition of evolutionary stability in the adaptive dynamics literature is *local* (*e.g.,* [57]), so the conditions for stability can be weaker, which is what our analysis revealed for the public goods games that we considered: cooperation is *locally* evolutionary and convergently stable *no matter what*. Note also that the notion of global ES is more relevant than the local ES in cases when the deviation of mutant strategies from the resident one are not small, *e.g.,* in the case of flexible decision-making (rather than genetically predetermined behaviour) [146].

The reason that our adaptive dynamics analysis did not detect contributing nothing ("defection") as an ESS is an artifact of the analysis method's focus on evolutionarily singular points (see appendix 3.B). Numerical simulations based on adaptive dynamics (*e.g.,* those done in [55, 155]) are not subject to this constraint.

Compared with the static theory, the adaptive dynamics framework has the advantage of being able to describe evolutionary *dynamics*, even far from ESSs. However, studying the dynamics is possible only if a particular fitness function is adopted. In this paper, we

focused on a general setting, without restricting attention to a particular benefit function $f(\eta)$, so that the inferences we make are as broad as possible.

With the adaptive dynamics framework, it is possible to describe and investigate the evolution of a dimorphic population (as opposed to a single mutant in an otherwise uniform resident population) but, again, only if a specific fitness function is chosen. Our third theorem (theorem 3.5.4), based on static theory, successfully considers manifestly dimorphic populations in order to broaden the scope of the stability results to include potential effects of invasion by a significant proportion of mutants (which is applicable to a number of biological situations). We find that cooperation is locally evolutionarily and convergently stable in a much stronger sense than typically considered: when it is stable to invasion by a single mutant, H^* is actually selected for no matter how large a proportion of the population is playing the mutant strategy.

Throughout this paper, we have retained the standard assumption that the underlying population is infinite. An infinite population size is often justified in the adaptive dynamics literature on the grounds that small populations are unlikely to persist due to demographic stochasticity [58, §2.1]. Of course, evolutionary stability predictions might differ in finite populations [156], a possibility that we will explore in further work.

Other possible complications that are not accounted for in our present analyses are the effects of a structured (*i.e.,* not well-mixed) population [22, 173], relatedness among some or all agents in the population [138, 169], assymetry or variability among individuals, due to differences in age, sex, resources, abilities or costs [21, 114, 174, 175], as well as inter- and intra-group competition affecting the division of resources [117, 168]. Furthermore, in the class of games we have studied, individual agents choose their level of contribution independently and without knowledge of other agents' contributions. However, it is possible that agents choose their contributions in sequence, or negotiate their levels of contributions [175, 176, 177].

Funding

We were supported by NSERC (DE), the Ontario Trillium Foundation (CM) and the department of Mathematics and Statistics at McMaster University (CM).

Conflict of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We are grateful to Sigal Balshine, Pat Barclay, Ben Bolker, Jonathan Dushoff, Paul Higgs, Rufus Johnstone and Danny Krupp for valuable discussions and comments.

Appendix

3.A Appendix: Motivation for assumption A3 (existence of $\eta_{\max}$)

In this appendix, we motivate assumption A3 by showing that if the focal agent's fitness is defined by equation (3.5) and f is continuous, the following two statements are equivalent:

S1 For any fixed non-focal agent's mean contribution H there exists $h^\dagger(H) \geq 0$ such that the focal agent's fitness $W(h, H)$ decreases with its contribution h for all $h > h^\dagger(H)$.

S2 There exists $\eta_{\max} \geq 0$ such that $f(\eta) - \eta$ decreases with η for all $\eta > \eta_{\max}$.

To gain some intuition, we first suppose f is a differentiable function of η (in §3.A.1), and then give the general proof (in §3.A.2)

3.A.1 If f is differentiable

Suppose that f is differentiable. Then, by the chain rule and equation (3.3), S1 implies that

$$\begin{aligned}\frac{\partial W}{\partial h} &= \left(f'(\eta)\frac{\partial \eta}{\partial h}\right)\Big|_{\eta=\eta(h,H)} - 1 = f'(\eta)|_{\eta=\eta(h,H)} - 1 \\ &= \frac{\mathrm{d}}{\mathrm{d}\eta}\Big(f(\eta) - \eta\Big)\Big|_{\eta=\eta(h,H)} . \end{aligned} \tag{3.51}$$

Consequently, if $W(h, H)$ decreases with h for $h > h^\dagger(H)$ then $f(\eta) - \eta$ decreases with η for all $\eta > \eta(h^\dagger(H), H)$. Letting

$$\eta_{\max} = \min_{H\geq 0} \eta(h^\dagger(H), H)\,, \tag{3.52}$$

$f(\eta) - \eta$ decreases for $\eta > \eta_{\max}$. Thus, S1 implies S2.

Conversely, if there exists $\eta_{\max} \geq 0$ such that $f(\eta) - \eta$ decreases for $\eta > \eta_{\max}$, then

letting $h^\dagger(H) = \eta_{\max} - (n-1)H$, we see that $\eta(h,H) > \eta_{\max}$ *iff* $h > h^\dagger(H)$. It then follows from equation (3.51) that $W(h,H)$ decreases with h for $h > h^\dagger(H)$, so S2 implies S1.

3.A.2 General case

Suppose S1 holds. Noting that

$$\begin{aligned} h > h^\dagger(H) &\iff h + (n-1)H > h^\dagger(H) + (n-1)H \\ &\iff \eta(h,H) > \eta^\dagger(H)\,, \end{aligned} \tag{3.53}$$

where

$$\eta^\dagger(H) = \eta\big(h^\dagger(H), H\big)\,, \tag{3.54}$$

and rewriting equation (3.5) as

$$\begin{aligned} W(h,H) &= f\big(h + (n-1)\,H\big) - h \\ &= f\big(\eta\,(h,H)\,\big) - \big[\eta\big(h,H\big) - (n-1)H\big] \\ &= f(\eta) - \eta + (n-1)H\,, \end{aligned} \tag{3.55}$$

we see that S1 is equivalent to the assumption that for fixed H there exists $\eta^\dagger(H) \geq 0$ such that $f(\eta) - \eta$ decreases for $\eta > \eta^\dagger(H)$. We define $\eta_{\max}$ to be the minimal such (non-negative) total good. Because η can vary independently of H, it follows that $f(\eta) - \eta$ decreases for all $\eta > \eta_{\max}$, so S1 implies S2.

Conversely, if S2 is true, then equations (3.55) and (3.53) imply that $W(h,H)$ decreases for all $h > h^\dagger(H) = \eta_{\max} - (n-1)H$. Thus, S2 implies S1.

3.B Appendix: Boundary ESSs need not be singular strategies

In adaptive dynamics, evolutionarily singular points are singled out as candidate ESSs (*e.g.,* [55, 57]). These are points at which there is no directional selection, since the fitness gradient $D(H)$ vanishes.

However, when the evolving variable H is restricted to an interval (in our case $H \geq 0$), it is not necessary for the fitness gradient to vanish at an endpoint of this interval in order for it to be ES: as we have seen in theorem 3.5.1, for the class of models defined in §3.3, the endpoint $H = 0$ is globally evolutionarily stable whenever $f(\eta_{\max}) < \eta_{\max}$, but the fitness gradient is negative in a right-hand neighbourhood of the endpoint $H = 0$ (including at

$H = 0$). In fact, it is $D(H)$ being negative near $H = 0$ that ensures that $H = 0$ is both locally convergently and evolutionarily stable.

The source of this issue is that the restriction to the biologically meaningful interval $H \geq 0$ is not built into the dynamical model equation (4.32), in that solutions of equation (4.32) do not necessarily remain in this interval (because the fitness gradient at the left endpoint $H = 0$ points outside the interval, into $H < 0$).

Note also that this cannot be easily fixed by artificially setting $D(0) = 0$, because doing so will insert a discontinuity into the fitness gradient, and adaptive dynamics assumes that the fitness gradient is at least continuous, in order to ensure the existence of solutions of equation (4.32) (see [178]) and in order to perform the local analysis leading to table 4.A.1.

We conclude that when using adaptive dynamics to model a trait that is restricted to an interval (for biological reasons), points on the boundary of this interval may be ES, despite not being singular points. More care is thus required to examine the dynamics of such models near boundary points.

3.C Appendix: The assumption that contribution is measured in units of fitness cost, $c(h) = h$

In this appendix, we comment on the biological interpretation of our assumption that the contribution of the focal agent is measured in units of the fitness cost it incurs, $c\,(h, H) = h$ (equation (3.2)).

Suppose, as before, that the population is engaged in an n-player public goods game, and let $h_1\,, \ldots, h_n$ be the contributions of all the members of the focal agent's group, including the focal agent (for example, if the index of the focal agent is $i = 1$, then $h = h_1$).

Thus, substituting

$$\eta(h, H) = \sum_{i=1}^{n} h_i \tag{3.56}$$

in equation (3.4), we have

$$b\,(h, H) = f\big(\eta(h, H)\big) = f\left(\sum_{i=1}^{n} h_i\right)\,. \tag{3.57}$$

However, we relax our assumption in equation (3.2) and instead only assume that

$$c(h, H) = c(h)\,, \tag{3.58}$$

so that the fitness cost incurred by the focal agent is independent of the contributions of the

other members in its group.

The focal agent's fitness is then

$$W(h, H) = f\left(\sum_{i=1}^{n} h_i\right) - c(h)\,. \tag{3.59}$$

For $1 \leq i \leq n$, let $C_i = c(h_i)$, and $C = c(h)$ be the costs incurred by the n members of the focal agent's group, and the focal agent (respectively). Suppose that the cost function is one-to-one, so that there exists a left-inverse function $k(\cdot)$ satisfying $k\big(c(h)\big) = h$ and $k\big(c(h_i)\big) = h_i$ for all $1 \leq i \leq n$. Then, equation (3.59) becomes

$$W(h, H) = f\left(\sum_{i=1}^{n} k(C_i)\right) - C\,. \tag{3.60}$$

The benefit to the focal agent, $f\left(\sum_{i=1}^{n} k(C_i)\right)$ is then generally not a function of the sum of the group members' fitness costs, $\sum_{i=1}^{n} C_i$.

By assuming that contributions to the public good are expressed in units of fitness cost (*i.e.,* $c(h) = h$, as in equation (3.2)), we implicitly assumed that fitness itself is the public good. Expressed in more biological terms, we are assuming that reproductive costs are effectively transferable: each individual in a group obtains a fitness benefit $f(\eta)$ regardless of how the associated costs (which sum to η) are distributed among the group members; for example, the fitness benefit is the same if the focal agent contributes the entire cost ($h = \eta$), or if the cost is distributed equally among group members ($h_i = \eta/n$ for each i).

Chapter 4

Evolutionarily stability in continuous public goods games in finite populations

Chai Molina and David J. D. Earn

4.1 Abstract

The evolution of cooperation is frequently investigated using public goods games. A classic example is the n-player snowdrift game, in which each player incurs a cost from contributing to a common good but benefits from the pooled contributions of all group members. Such games arise in many biological contexts, from bacterial communities to human societies. With a continuum of contribution strategies (*e.g.,* time devoted to a task benefiting the community), analyses to date have typically assumed—for mathematical convenience—that groups are drawn from an infinite population. Here, we rigorously analyze the continuous n-player snowdrift game in finite populations and compare the evolutionary outcomes with those in infinite populations. We show that evolutionarily stable strategies (ESSs) in infinite populations are always *un*stable when played in finite populations: selection favours invasion and fixation by less cooperative mutants. We demonstrate that in a large class of snowdrift games that always have a cooperative ESS in infinite populations, there may be no cooperative ESS in a finite population, even for arbitrarily large population size. We show that in such cases, not contributing is a globally convergently stable finite-population ESS, implying that apparent evolution of cooperation in such games is an artifact of the infinite population approximation. In addition, we find that in finite-population snowdrift games in which cooperation *can* evolve, a large population size is often required. Our results are robust to the underlying selection process.

Statement of significance

The ubiquity of cooperation in the living world contrasts with our understanding of evolution by natural selection, which favours traits promoting individual reproductive success. Theoretical studies of cooperation are often based on an idealized "public goods game" in which individuals contribute to a resource that is shared by all members of their group. For mathematical convenience, it is usually assumed that the population in which groups reside is infinite. We show that conclusions drawn from such infinite population models can be misleading. In particular, the possibility of cooperation evolving may require a large population size. Moreover, there are situations in which cooperation would persist if the population were infinite but cannot persist in *any* finite population.

4.2 Introduction

Self-interest is a fundamental component of the theory of evolution by natural selection [18, 179, 180]. The ubiquity of cooperative behaviour in the living world [122, 123, 181] is surprising to biologists and has motivated much theoretical work aiming to demystify its evolution [19, 20, 21, 22, 23].

Of particular interest in the study of cooperation are public goods games [54, 55, 108]. Public goods are commodities that are *non-rival*[1] (*i.e.,* the amount of good available for consumption by an individual is independent of others' consumption) and *non-excludable* (*i.e.,* it is impossible to exclude others from sharing in the resulting benefit) [182]. Thus, the cost of contribution to the public good is personal, but the benefit is shared. Many biological and social phenomena involve public goods: microbial evolution [34, 35], tumor growth [110], the evolution of virulence [111], host manipulation by parasites, [112], rhizobia-legume mutualism [140], cooperative nesting and brood care [113, 114, 115, 116], the evolution of eusociality [117], fisheries management [118], family economics [119], voluntary organizations (*e.g.,* neighbourhood watch [183]) and vaccination ([36, 37] and chapter 2), to name a few. See references [125, 126] for recent reviews.

Many evolutionary games assume—for mathematical convenience—that populations are infinitely large (*e.g.,* [55, 57, 141, 148, 151, 184, 185]). This assumption is sometimes justified on the grounds that "[p]opulations which stay numerically small quickly go extinct by chance fluctuations" [58, §2.1]. Of course, all real populations are finite, and important differences in evolutionary dynamics between finite and infinite populations have been demonstrated [81, 129, 186, 187, 188, 189, 190]. In spite of the technical challenges of working with finite populations, some exact analytical results have been obtained for two-player games with discrete strategy sets [81, 129, 186, 189, 191]. How-

[1] Some biological and social examples involve rival goods, of which consumption by one individual diminishes the amount available to others, *e.g.,* [34, 116].

ever, most existing finite-population results rely on approximation methods and simulations [129, 153, 188, 192, 193, 194, 195]. Notably, almost all finite-population results involve discrete strategy sets (such as when individuals must choose between making a fixed positive contribution to a public good, or nothing at all). Yet, evolutionary games involving continuous strategy sets (*e.g.,* allocating time or effort to a communal task) are both widely applicable and extensively studied [136]. Moreover, to our knowledge, all existing results for finite populations depend on a choice of selection process (*e.g.,* Moran or Wright-Fisher [61, 62]).

Here, we focus on the continuous n-player ***snowdrift game*** [55], previously studied in infinite populations [55, 146, 147, 151, 172] and finite populations (the latter using approximations and individual-based simulations) [153, 188, 196, 197]. In this game, groups of n self-interested agents select their contribution to a public good shared among the group. A focal agent contributing h incurs a cost $C(h)$ that depends only on its contribution, whereas its benefit $f(\eta)$ depends on the total good contributed by the group, η. The focal agent's payoff is then $f(\eta) - C(h)$.

We show that strategies predicted to be evolutionarily stable (ES; see appendix 4.A.1) in infinite populations (and in fact all singular strategies; see appendix 4.A.2) are not ES in finite populations: selection always favours both invasion and replacement of a monomorphic population playing such a strategy by a slightly less cooperative one. Consequently, we propose an extension of the concept of evolutionarily singular strategies to finite populations. We then find exact analytical conditions for selection opposing invasion, convergent stability [158] (see appendix 4.A.1) and evolutionary branching in finite populations. These conditions are different from the corresponding ones obtained for infinite populations, but approach them as the population size is increased (while keeping the size of interacting groups small relative to the population size). These results apply generally to any snowdrift game played in (haploid) finite populations, for any selection process (*e.g.,* Moran or Wright-Fisher [61, 62]). The extension of singular points to finite populations, as well as our approach in qualitatively characterizing these points can easily be applied to many other multi-player games.

The quantitative difference in singular strategies between finite and infinite populations seems to have been largely overlooked [188, 193, 194] , possibly because the discrepancy is often small. However, applying our results to a class of snowdrift games previously analyzed in infinite populations in chapter 3, we find conditions under which there is no cooperative strategy at which selection opposes invasion in finite populations, despite there being a cooperative ES strategy if the population is infinite. This qualitative difference between finite and infinite populations can persist for arbitrarily large population sizes, and is independent of the selection process (*e.g.,* the Moran or Wright-Fisher processes [61]). To our knowledge, there are no other examples in the literature for a qualitative difference in invasion dynamics between finite and infinite populations that persists for arbitrarily large populations, or for any selection process; all other such differences demonstrated concern

fixation of strategies, and are restricted to a particular selection process (most often the Moran process).

Our results are summarized in § 4.3 and proved in § 4.6. The discrepancy between evolutionary outcomes in finite and infinite populations, and the reasons for it, are discussed in §4.4. As an example, these differences are precisely identified for a subclass of snowdrift games in §4.4.2, where we find conditions under which qualitative differences exist between evolutionary outcomes in finite and infinite populations.

4.3 Results

We wish to analyze the continuous n-player snowdrift game in a finite population of constant size N. To do so, first recall that the definition of evolutionary stability in finite populations (ESS_N) must account for the fact that selection can favour fixation of a mutant strategy, even if selection opposes its invasion [186] (which is not the case in infinite populations). Thus, the standard definition of evolutionary stability in a finite population (see definition 5.6.3) requires that selection opposes both invasion by mutants, and fixation of mutant strategies.

The fixation probabilities of mutant strategies depend on the stochastic process that specifies how the variability in the game payoff generates changes in the frequencies of strategies in the population, which we call the ***selection process*** (see chapter 5). The Moran or Wright-Fisher processes [61, 62], or their frequency-dependent analogues [81, 186] are common choices for this process. However, in order to maintain generality, we avoid specifying a population-genetic process throughout this article.

In analyses of the continuous n-player snowdrift game in infinite populations (based either on adaptive dynamics [55] or "static" evolutionary game theory, as in chapter 3), ESSs must be evolutionarily singular strategies (see appendix 4.A.2). Theorems 4.3.1 and 4.3.2 (below) are concerned with the evolutionary *instability* of these strategies in a finite population: Roughly speaking, theorem 4.3.1 shows that strategies that are evolutionarily singular in an infinite population cannot be ESS_Ns (in particular, a single mutant can invade if the population is finite). Theorem 4.3.2 shows that when strategies that are evolutionarily singular in infinite populations are played by any number of residents in a finite population, they are selected to be ousted (that is, replaced entirely) by sufficiently similar strategies that are less cooperative. Theorem 4.3.2 is not simply a stronger version of theorem 4.3.1. The two results have distinct hypotheses and distinct conclusions. Theorem 4.3.1 is easier to connect with the results available for infinite populations, while theorem 4.3.2 provides more information about games in finite populations.

Theorem 4.3.1 (Strategies that are singular in infinite populations are not ESS_Ns)**.** *Consider an evolving, finite population of N agents, composed of G groups of size $n > 1$ (so $N =$*

Gn). Suppose fitness is determined by the payoff from a public goods game played in each group. Denote a focal agent's contribution to the total public good by h, the mean contribution of the $n-1$ other agents in the focal agent's group by H, and the total good contributed in the focal agent's group by

$$\eta(h,H) = h + (n-1)H\,. \tag{4.1}$$

Let $C(h)$ be the cost incurred by the focal agent for its contribution h, and suppose that the benefit it obtains is a function of the total good contributed in its group, $f\big(\eta(h,H)\big)$. Thus, the focal agent's fitness is

$$W(h,H) = f\big(h + (n-1)H\big) - C(h)\,. \tag{4.2}$$

Let H_{s} be an evolutionarily singular strategy in the infinite population analogue of the game above, and $\eta_{\mathrm{s}} = nH_{\mathrm{s}}$ be the total good when all group members play H_{s}.

Assume that the cost and benefit functions $f(\eta)$ and $C(h)$ are continuously differentiable in some neighbourhoods of $\eta = \eta_{\mathrm{s}}$ and $h = H_{\mathrm{s}}$, respectively, and that $C'(H_{\mathrm{s}}) \neq 0$. Suppose that a population of agents playing H_{s} is invaded by a mutant strategy $h \neq H_{\mathrm{s}}$, and that groups playing the public goods game can contain at most one mutant.

- *if $C'(H_{\mathrm{s}}) > 0$, then, if $h < H_{\mathrm{s}}$ is sufficiently close to H_{s}, mutants (playing h) obtain a higher fitness than residents (playing H_{s}).*
- *if $C'(H_{\mathrm{s}}) < 0$, then, if $h > H_{\mathrm{s}}$ is sufficiently close to H_{s}, mutants (playing h) obtain a higher fitness than residents (playing H_{s}).*

Consequently, H_{s} is not evolutionarily stable. Moreover, these conclusions apply in the infinite population limit as well, in the following sense: if an infinite population playing a singular strategy H_{s} is invaded by a finite proportion *$\epsilon > 0$ of mutants, and groups playing the game contain at most one mutant, then if $C'(H_{\mathrm{s}}) > 0$ ($C'(H_{\mathrm{s}}) < 0$) mutants playing a strategy $h < H_{\mathrm{s}}$ (respectively, $h > H_{\mathrm{s}}$) sufficiently close to H obtain a higher fitness than the residents.*

The reason that theorem 4.3.1 can be applied to both finite and infinite populations is that its proof is independent of whether groups are formed by sampling from the population with or without replacement. By contrast, theorem 4.3.2 below applies only to finite populations, and its results depend on individuals being sampled from the population without replacement when a group is formed. The role of the group sampling procedure in determining evolutionary stability is discussed further in §4.4.1.

Theorem 4.3.2 (Selection favours replacement of strategies that are singular in infinite populations)**.** *Suppose that the hypotheses of theorem 4.3.1 are satisfied, except that groups are randomly sampled from the population and can therefore contain $0 \leq k \leq n$ mutants (rather than at most one). Then, for any K such that $1 \leq K < N$, if K agents are mutants*

playing $h \neq H_{\mathrm{s}}$, $N-K$ agents are residents playing H_{s},

- *if $C'(H_{\mathrm{s}}) > 0$ and $h < H_{\mathrm{s}}$ is sufficiently close to H_{s}, then the mean fitness of mutants playing h is greater than the mean fitness of residents playing H_{s}.*
- *if $C'(H_{\mathrm{s}}) < 0$ and $h > H_{\mathrm{s}}$ is sufficiently close to H_{s}, then the mean fitness of mutants playing h is greater than the mean fitness of residents playing H_{s}.*

Without specifying a particular selection process, it is impossible to calculate the probability at which a mutant strategy fixes. Yet, biological intuition suggests that if the mutant's fitness is higher than the residents' for any number of mutants $1 \leq K < N$, then selection does indeed favour fixation of the mutant strategy. Theorem 4.3.2 and corollary 5.6.5 imply that this intuition is, in fact, correct for *any* selection process:

Corollary 4.3.3. *Suppose that the hypotheses of theorem 4.3.2 are satisfied. If $C'(H_{\mathrm{s}}) > 0$ ($C'(H_{\mathrm{s}}) < 0$), then for any selection process, selection favours fixation of mutant strategies $h < H_{\mathrm{s}}$ (respectively, $h > H_{\mathrm{s}}$) sufficiently close to the resident singular strategy H_{s}.*

Since selection favours invasion of strategies that are evolutionarily singular in infinite populations (according to the standard definition given in appendix 4.A.2), a different notion of singular strategy must be appropriate in finite populations. Recall that in infinite populations, directional selection vanishes near singular strategies [55, 57]. In contrast, in a finite population playing an infinite-population singular strategy, directional selection does *not* vanish, as a result of the presence of a single invader affecting the mean resident fitness (see equation (4.19) below). We therefore take the qualitative condition of *directional selection vanishing* as the definition of a singular strategy, which applies to both finite and infinite populations. To make this precise, we first introduce notation for the mean fitness difference between mutants and residents.

Definition 4.3.4 (Mean fitness difference)**.** *Consider a finite or infinite population playing the public goods game described in theorem 4.3.1. Suppose there are two strategies in the population, a resident strategy H and a mutant strategy h. We denote the* ***mean fitness difference*** *between mutants and residents by*

$$\delta\overline{W}(h,H) = \overline{W}_{\mathrm{m}}(h,H) - \overline{W}_{\mathrm{r}}(h,H)\,. \tag{4.3}$$

Note that the mean fitness difference depends on the number or proportion playing the mutant strategy.

Definition 4.3.5 (Singular strategy)**.** *We say that H is an* ***evolutionarily singular strategy*** *if, when a single mutant playing h invades a population of residents playing H, directional selection vanishes as the mutant strategy h approaches H, that is,*

$$\partial_h \delta\overline{W}(h,H)|_{h=H} = 0\,. \tag{4.4}$$

Remark 4.3.6. *Note that while equation (4.4) was derived in references [197, 198], their derivations employ classical adaptive dynamics [57, 58] and, as such, assume infinite populations. Thus, while obtaining the correct condition for a strategy being singular, these derivations were flawed, because the $N \to \infty$ limit was taken inconsistently.*

We now apply this definition to the continuous n-player snowdrift game, to obtain a characterization of singular strategies:

Lemma 4.3.7 (Selection opposes invasion)**.** *Suppose that a population of size N is playing the public goods game described in theorem 4.3.1.Then, selection opposes invasion of a population playing $H = \hat{H}$ by sufficiently similar mutant strategies $h \neq \hat{H}$* iff *$h = \hat{H}$ is a local maximum of the mean fitness difference $\delta\overline{W}(h, \hat{H})$.*

As an immediate corollary of lemma 4.3.7, we obtain

Corollary 4.3.8 (Cooperative strategies at which selection opposes invasion are singular)**.** *Suppose that a population of size N is playing the public goods game described in theorem 4.3.1. If selection opposes invasion of a population playing a cooperative strategy $H = \hat{H} > 0$ by mutant strategies $h \neq \hat{H}$, and the cost and benefit functions $C(h)$ and $f(\eta)$ are differentiable in a neighbourhood of $\hat{H}$ and $n\hat{H}$, respectively, then $\hat{H}$ is a singular strategy (according to definition 4.3.5).*

Sufficient conditions for selection opposing invasion and convergent stability in a finite population are given in the following theorem:

Theorem 4.3.9 (Conditions for selection opposing invasion and convergent stability)**.** *Suppose that a population of size N is playing the public goods game described in theorem 4.3.1, with twice-differentiable cost and benefit functions, $C(h)$ and $f(\eta)$.*

If $\hat{H} > 0$: *Selection opposes invasion of a population playing $H = \hat{H}$ by a sufficiently similar mutant strategy $h \neq \hat{H}$ if the mean fitness difference $\delta\overline{W}$ satisfies two conditions:*

$$\textit{(i)} \qquad \partial_h \delta\overline{W}(h, \hat{H})|_{h=\hat{H}} = \frac{N-n}{N-1} f'\big(n\hat{H}\big) - C'(\hat{H}) = 0\,, \tag{4.5}$$

$$\textit{(ii)} \qquad \partial_h^2 \delta\overline{W}(h, \hat{H})|_{h=\hat{H}} = \frac{N-n}{N-1} f''\big(n\hat{H}\big) - C''(\hat{H}) < 0\,. \tag{4.6}$$

Condition (i) ensures that $\hat{H}$ is a singular strategy according to definition 4.3.5, while condition (ii) ensures that $h = \hat{H}$ is a local maximum point of $\delta\overline{W}(h, \hat{H})$.
$\hat{H}$ is convergently stable if condition (i) holds, together with

$$n\frac{N-n}{N-1} f''\big(n\hat{H}\big) - C''(\hat{H}) < 0\,. \tag{4.7}$$

If $\hat{H} = 0$: *Selection opposes invasion if the mean fitness difference $\delta\overline{W}$ satisfies conditions (4.5) and (4.6), or if*

$$\partial_h \delta\overline{W}(h,0)|_{h=0} = \frac{N-n}{N-1} f'\big(0\big) - C'(0) < 0\,. \tag{4.8}$$

$\hat{H} = 0$ is convergently stable if, for sufficiently small $H > 0$,

$$\frac{N-n}{N-1} f'\big(nH\big) - C'(H) < 0\,. \tag{4.9}$$

Remark 4.3.10. *In the infinite population limit ($N \to \infty$), if the group size becomes negligible compared to the population size ($n/N \to 0$), then conditions (4.5) and (4.6) approach the ESS conditions derived from classical adaptive dynamics [55]. A recent extension of adaptive dynamics can also be used to construct a dynamical system to analyze the public goods game described in theorem 4.3.1 played in structured populations [79]. When the "updating rule" is either the Moran or Wright-Fisher process and the population is well-mixed, conditions (4.5) and (4.6) also characterize attractors of this dynamical system.*

In appendix 4.D, we give a sufficient condition for evolutionary stability (ESS_N) in finite population n-player snowdrift games, that is independent of the selection process.

As in the infinite population case, a finite-population singular strategy $\hat{H}$ will be an ***evolutionary branching point*** [55] if it is convergently stable, but selection favours invasion by sufficiently similar strategies; the latter occurs if $h = \hat{H}$ is a local minimum of the mean fitness difference $\delta\overline{W}(h,\hat{H})$. A sufficient condition for a singular stragtegy $h = \hat{H}$ to be a local minimum of $\delta\overline{W}(h,\hat{H})$ is obtained by reversing condition (4.6), which combined with convergent stability (condition (4.7)) gives

$$\frac{N-n}{N-1} f''\big(n\hat{H}\big) < C(\hat{H}) < n\frac{N-n}{N-1} f''\big(n\hat{H}\big)\,. \tag{4.10}$$

Remark 4.3.11. *Note that all of our results remain true also if fitness is an increasing affine function (*i.e., *of the form $\phi(x) = w \cdot x + k$, with $w > 0$) of the game payoff, as this would simply add a positive scaling factor to all of our conditions, which involve only derivatives of the fitness. In particular, these results are independent of the slope of this affine function (w above), which is typically referred to as the intensity of selection [156] (and taken to be $0 \leq w \leq 1$).*

Remark 4.3.12. *Observe that if the n-player snowdrift game described above is played repeatedly between reproductive events, and the agents' fitnesses are calculated using their mean payoffs, the expected fitnesses for residents and mutants are unchanged, and hence, all of our results above remain true. In this scenario, the stipulation that the population is divided into G groups of n agents is unnecessary.*

4.4 Discussion

Comparing theorems 4.3.1, 4.3.2 and 4.3.9 shows that public goods games yield different outcomes when played in finite *vs.* infinite populations. Inferences made from analyses employing an infinite population approximation should be interpreted carefully and cautiously:

- Theorem 4.3.9 gives conditions for selection opposing invasion that are different from those obtained for infinite populations ([55] and chapter 3).
- Theorems 4.3.1 and 4.3.2 show that strategies expected to be ESSs based on infinite-population models cannot be evolutionarily stable (ES) in finite populations and are, in fact, selected to be replaced by sufficiently similar strategies.

The contrast between the predicted evolutionary outcomes in finite and infinite populations becomes even more striking when focusing on a slightly less general version of the n-player snowdrift game, introduced in chapter 3, which assumes the following:

(a) The cost to the focal agent of a contribution h is measured in units of its impact on this agent's fitness, that is, $C(h) = h$.

(b) The benefit to group members in which the total group contribution is $\eta \geq 0$ is a continuous function $f(\eta)$.

(c) There exist $\eta_{\min} \geq 0$ and $\eta_{\max} > \eta_{\min}$ such that $f(\eta) - \eta$ increases for $\eta_{\min} < \eta < \eta_{\max}$ and decreases for $\eta < \eta_{\min}$ and $\eta > \eta_{\max}$.

(d) $f(\eta_{\max}) < \eta_{\max}$ and

$$f(\eta_{\max}) - f\Big(\frac{n-1}{n}\eta_{\max}\Big) > \frac{\eta_{\max}}{n}\,. \tag{4.11}$$

In order to use theorem 4.3.9 and to facilitate comparison with the results from adaptive dynamics [55], we also assume f is twice-differentiable in a neighbourhood of $\eta_{\max}$.

When played in an infinite population, assumptions (a)-(d) ensure the existence of a cooperative global ESS, $H^*_\infty = \eta_{\max}/n$, where $\eta_{\max}$ is the point at which $f(\eta) - \eta$ is maximal (theorem 3.5.1)[2] [3]. As seen in Theorem 3.5.3, if the benefit function f is differentiable at

[2] Note that the case $f(\eta_{\max}) \geq \eta_{\max}$ is excluded despite it also implying the existence of a cooperative ESS; this case is uninteresting as a public goods game, since then, the focal agent benefits from contributing a nonzero amount ($h = \eta_{\max}$), even if no-one else contributes.

[3] In terms of the ESS, H^*_∞, condition (4.11) can be rewritten as

$$f\big(nH^*_\infty\big) - f\big((n-1)H^*_\infty\big) > H^*_\infty\,, \tag{4.12}$$

which can be interpreted as: when all nonfocal agents contribute H^*_∞, the incremental benefit to the focal agent of also contributing H^*_∞ exceeds its cost.

$\eta_{\max}$, then H^*_∞ must be an evolutionarily singular strategy, that is, the fitness gradient must vanish there ($D(H^*_\infty) = f'(nH^*_\infty) - 1 = 0$, where the fitness gradient $D(H)$ is defined as in appendix 4.A.2). However, when the game defined in chapter 3 is played in a finite population, by theorem 4.3.1, H^*_∞ is *not* an ESS, for any finite population size, N.

Moreover, in an infinite population playing the ESS H^*_∞, if a proportion ϵ of mutants invade and play a different strategy $h \neq H^*_\infty$ that is sufficiently close to H^*_∞, then h is selected against (*i.e.,* the mutants' fitness is lower than the residents'), regardless of ϵ (theorem 3.5.4). Yet, in a finite population, theorem 4.3.2 extends theorem 4.3.1 to *any* number of mutants arising in the population and to how many mutants can be in a group: if in a population of N agents, $N - K$ play H^*_∞ and K play an alternative strategy h, then for any K, strategies $h < H^*_\infty$ sufficiently close to H^*_∞ yield a higher fitness than H^*_∞. Moreover, corollary 4.3.3 states that strategies $h < H^*_\infty$ sufficiently close to H^*_∞ are selected to entirely replace the resident strategy H^*_∞.

These differences motivate us to explore how evolutionary outcomes of continuous snowdrift games differ between finite and infinite populations, both qualitatively and quantitatively, and to understand the underlying reasons for these different outcomes. We address these issues in §§4.4.1 and 4.4.2.

4.4.1 Why evolutionary stability in infinite populations does not imply resistance to invasion in finite populations

Table 4.1 summarizes the differences between theorems 3.5.1 and 3.5.4, and theorems 4.3.1 and 4.3.2 of this paper, applied to the infinite-population ESS H^*_∞. In this section, we explain how these differences arise.

Theorem 3.5.1 states that when an infinite population playing H^*_∞ is invaded by a single mutant playing a different strategy, $h \neq H^*_\infty$, the mean resident fitness (which is unaffected by the mutant) is higher than the mean mutant fitness and H^*_∞ is ES. By contrast, when interacting groups contain at most one mutant, theorem 4.3.1 shows that an infinite population playing H^*_∞ is susceptible to invasion by a proportion $0 < \epsilon < 1$ of less cooperative mutants. This is because resident–mutant interactions decrease the mean resident fitness, but the mean mutant fitness is unchanged (because mutants interact only with residents). Hence, infinitely diluting of the mutants on the residents is essential to the stability of the infinite population ESS H^*_∞.

Now consider the invasion of infinite and finite populations (respectively) by a non-negligible number of mutants, when groups are sampled randomly from the population. If an infinite population of residents playing H^*_∞ is invaded by a non-negligible proportion $0 < \epsilon < 1$ of mutants playing a strategy $h \neq H^*_\infty$ sufficiently close to H^*_∞, the mean resident fitness is still higher than the mean mutant fitness (Theorem 3.5.4). Yet, in a finite population, theorem 4.3.2 shows that a slightly less cooperative mutant strategy $h < H^*_\infty$

yields a higher mean fitness than the resident strategy H^*_∞ (for any number of the mutants in the population $1 \leq K \leq N-1$). It appears that another process must be at work that causes the infinite-population ESS to be unstable when played in a finite population (in addition to the infinite dilution of the effects of mutants on other agents).

To identify this additional process destabilizing the infinite-population ESS H^*_∞, we make two more comparisions of scenarios in which populations playing H^*_∞ are invaded by a proportion $0 < \epsilon < 1$ of mutants playing a sufficiently similar, less cooperative strategy $h < H^*_\infty$ (see table 4.1:

- Theorem 3.5.4 *vs.* theorem 4.3.1: mutants cannot invade when groups are randomly sampled from the population *vs.* mutants can invade when groups contain at most one mutant (and thus mutants do not interact). Consequently, mutant–mutant interactions adversely affect the mean mutant fitness.
- Theorem 3.5.4 *vs.* theorem 4.3.2: mutants cannot invade an infinite population *vs.* can invade a finite population. The difference arises because when a finite group is sampled from a population, if the population is infinite then the relative frequencies of the mutants and residents are not changed, whereas they *are* changed if the population is finite. Consequently, in an infinite population, on average, residents interact with fewer mutants and mutants interact with more mutants, compared to a finite population with the same proportion of mutants (see appendix 4.G). Thus when mutants are less cooperative than residents, mutants are better off in a finite population, and residents are better off in an infinite one (because the focal agent's fitness decreases with the number of mutants in its group).

Approximating large populations by infinite ones may generate misleading conclusions, by neglecting or incorrectly estimating the effects of mutants on the agents interacting with them; this can be caused either by "infinitely diluting" the effects of a finite number of mutants, or by the difference in the mean number of nonfocal mutants in a group between finite and infinite populations (which is in turn caused by "diluting" the effect of sampling on the population composition).

4.4.2 Qualitative and quantitative differences between finite and infinite populations

We have seen that the evolutionary dynamics of the level of contribution in continuous snowdrift games in a finite population differ from those predicted by models based on an infinite population. Since all real populations are finite, we wish to evaluate how conclusions based on an infinite-population analysis of the n-player snowdrift game (*e.g.,* [55, 138] and chapter 3) might be affected.

In appendix 4.B, we address this issue by analyzing the slightly more restricted snow-

Theorem	**Population size**	**Residents interact with mutants?**	**Mutants interact with mutants?**	**Mutants' mean fitness higher than residents'?**
Theorem 3.5.1	∞	No	No	No
Theorem 4.3.1	N or ∞	Yes	No	Yes
Theorem 3.5.4	∞	Yes	Yes	No
Theorem 4.3.2	N	Yes	Yes	Yes

Table 4.1: Comparison of stability results for the infinite population ESS, H^*_∞, in finite and infinite populations (note that theorems 4.3.1 and 4.3.2 apply to any strategy that is singular in an infinite population, but are here applied to ESSs).

drift game from chapter 3 described above. We show that if there is a cooperative ESS_N when the game is played in a finite population, then the infinite-population cooperative ESS is always larger than the finite population ESS_N. For a large enough number of groups there is always a cooperative ESS_N, $H^*_N > 0$. When the cooperative ESS_N, H^*_N, exists, it quickly approaches H^*_∞ (see figure 4.1, top panel).

More importantly, we show that in contrast to the prediction of the infinite-population analysis (whereby there is always a cooperative ESS), if

$$\max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) < 1 + \frac{n-1}{N-n}\,, \tag{4.13}$$

then there is no cooperative ESS_N; in this case, defection (*i.e.,* $H = 0$) is a globally convergently stable ESS_N.

This qualitative difference between the outcomes in finite and infinite populations can occur either for small or large population sizes, or indeed for any population size (see figure 4.1, bottom panel):

- If $\max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) < 2$, then for small enough population size (*e.g.,* $N = 2n$) there is no cooperative ESS_N.
- If, for a fixed number of groups G,

$$1 + \frac{1}{2(G-1)} < \max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) < 1 + \frac{1}{G-1}\,, \tag{4.14}$$

  then for low group sizes there is a cooperative ESS_N, but as group size n (and consequently population size N) is increased, there is a group size beyond which there is

no longer a cooperative $\mathrm{ESS_N}$ (even for arbitrarily large group and population sizes).

- If, for a fixed number of groups G,

$$\max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) < 1 + \frac{1}{2(G-1)}\,, \tag{4.15}$$

then there is no cooperative $\mathrm{ESS_N}$, for *any* population size N.

These results contrast earlier work, which found general agreement between adaptive dynamics and stochastic simulations of finite populations [199] and specific agreement between the finite- and infinite-population evolutionary dynamics of the n-player snowdrift game with discrete strategies [129][4]. It is worth emphasizing, moreover, that the existence or lack of a cooperative $\mathrm{ESS_N}$ in the snowdrift game considered here is entirely independent of the selection process.

Because of the discrepancy between the conditions for evolutionary branching in a finite population (equation (4.3)) and in an infinite population [55], there may in principle also be situations in which evolutionary branching is predicted in one, but not the other. We plan to investigate this additional potential qualitative difference between finite and infinite populations in future work.

4.5 Conclusion

We analyzed general continuous n-player snowdrift games in finite populations and showed that cooperative strategies that are evolutionarily stable in infinite populations are unstable in finite populations and are selected to be replaced by less cooperative strategies. These results motivated a revised definition of evolutionarily singular points. We obtained conditions for selection opposing invasion by sufficiently similar mutants, convergent stability, and evolutionary branching, at a singular point, as well as a sufficient condition for evolutionary stability. These conditions differ from their infinite-population analogues, but approach them as population size is increased. The rapid convergence of the conditions defining and classifying singular points in a finite population, to the corresponding infinite-population conditions, may explain why this difference between finite and infinite populations has been overlooked in recent studies that relied on adaptive dynamics to study continuous traits evolving in finite populations [188, 194].

We applied our finite-population analysis to a class of continuous n-player snowdrift games, which have a cooperative ESS when played in infinite populations (see § 4.4.2 and appendix 4.B). For games in this class, if there is a cooperative $\mathrm{ESS_N}$ for the finite-population game, it is always less cooperative than the infinite-population ESS. Further-

[4]Note, however, that [129] also found that defectors prevail when the group size approaches the population size, even in situations where cooperators and defectors can coexist in an infinite population.

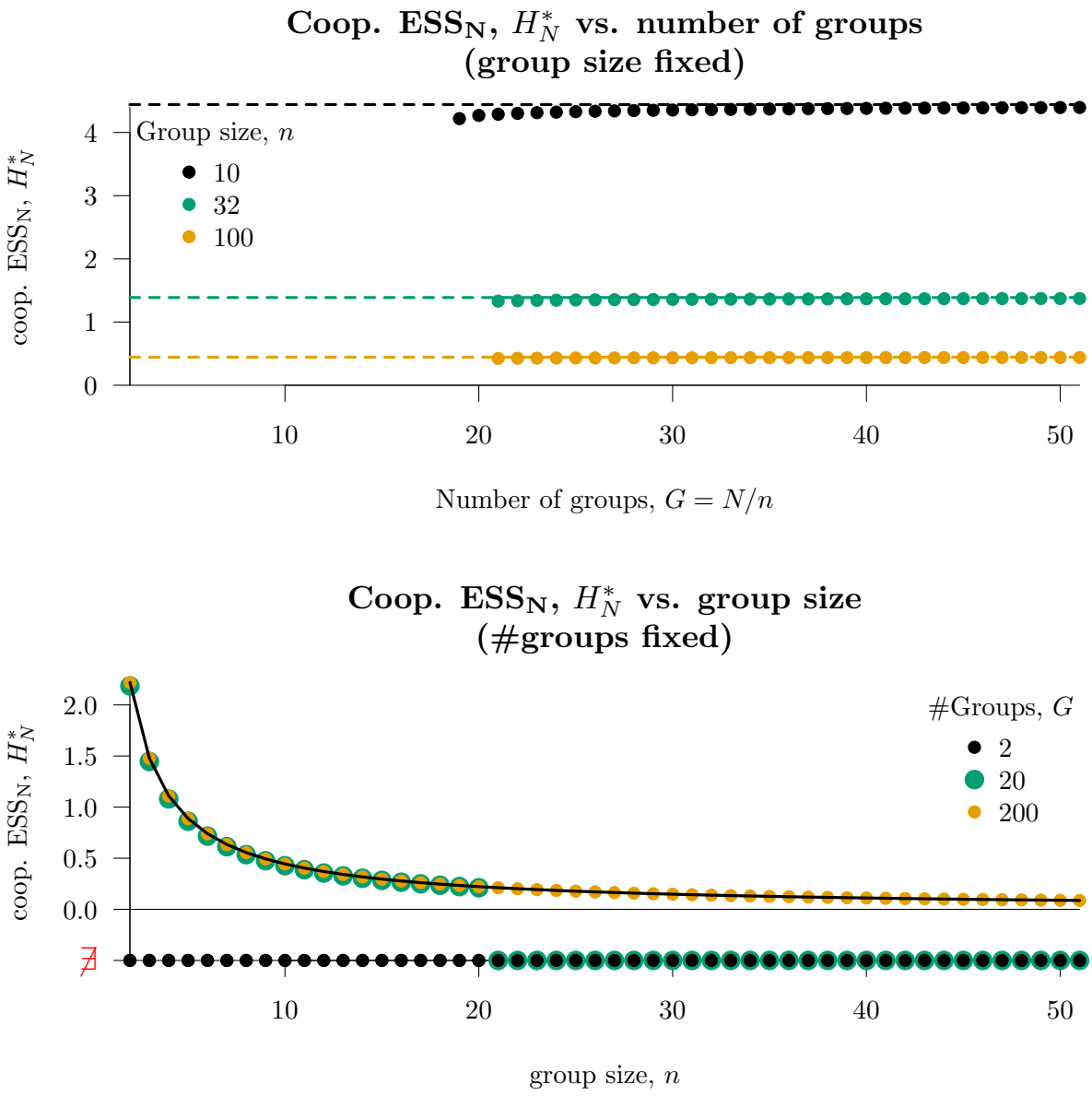

Figure 4.1: The finite-population ESS$_N$, H_N^*, as population size N is increased, for linear cost $C(h) = h$ and sigmoidal benefit $f(\eta) = a\big(\beta+\exp(\kappa-b\eta)\big)^{-1} - a\big(\beta+\exp(\kappa)\big)^{-1}$, with $a = 100$, $b = 0.2$, $\beta = 4.76$, $\kappa = 10$ (based on an example from [138]). Top panel: H_N^* is plotted against the number of groups, G, for three group sizes n. For each group size, a dashed line indicates the corresponding infinite-population ESS (H_∞^*). Bottom panel: H_N^* is plotted against the group size n for three different numbers of groups, G . The black curve is $H_\infty^* = \eta_{\max}/n$. When no cooperative ESS$_N$ exists, a point is plotted at the bottom of the figure beneath the horizontal axis ($\nexists$).

more, we identified conditions under which there is no cooperative ESS_N for such games when played in a finite population. Under these conditions, defection is a globally convergently stable ESS_N. Moreover, we found conditions under which no cooperative ESS_N exists for small population sizes, arbitrarily large population sizes, or any population size. Thus, the qualitative difference between the predictions of finite- and infinite-population models does not necessarily disappear as population size is increased. These results are independent of the selection process (*e.g.,* the Moran or Wright-Fisher processes).

In a broader context, our results indicate that conclusions drawn from models involving infinite populations may not extend to finite populations, despite previous work justifying the approximation of large populations by infinite ones (*e.g.,* [56]). In particular, adaptive dynamics has been extensively used in the study of evolutionary dynamics (*e.g.,* [55, 138, 155], as well as [200] and references therein), and relies on such an approximation [58]. Since all real populations are finite, our results indicate that some conclusions based on adaptive dynamics may not apply to realistic scenarios. This highlights the need to reevaluate the theoretical justification for approximating large populations by infinite ones, and to derive clear conditions for when such approximations are useful. We conjecture that in order to ensure that results based on infinite-population models apply qualitatively to their finite-population analogues, it is necessary that the size of an interacting group must be small compared to the population size ($n \ll N$).

4.6 Proofs of theorems

4.6.1 Proof of theorem 4.3.1

Consider a population of N agents playing $H_\text{s} = \eta_\text{s}/n$ invaded by a proportion $\epsilon = K/N > 0$ of mutants playing $h \neq H_\text{s}$, where interacting groups contain at most one mutant. This can happen if there is only one mutant in the population ($\epsilon = 1/N$) or because there are fewer mutants than there are interacting groups ($K < N/n = G$) and mutants can recognize and avoid one another. In this scenario, approximating group formation as sampling with replacement (using the binomial distribution, as done in theorem 3.5.4) is not appropriate.

There are K groups containing $n-1$ residents (that is, $K(n-1)$ residents in groups with one mutant) and $G-K$ groups containing n residents (that is, $N-Kn$ residents in groups with residents only). Then, the mean fitness of a resident is

$$\begin{aligned}\overline{W}_\text{r}(h) &= \frac{K(n-1)}{N-K} W(H_\text{s}, \frac{h+(n-2)H_\text{s}}{n-1}) + \frac{N-Kn}{N-K} W(H_\text{s}, H_\text{s}) \\ &= \frac{1}{1-\epsilon}\left(\epsilon(n-1)W(H_\text{s}, \frac{h+(n-2)H_\text{s}}{n-1}) + (1-\epsilon n)W(H_\text{s}, H_\text{s})\right), \end{aligned} \tag{4.16}$$

with $\epsilon = K/N$, and the fitness of a mutant is $W(h, H)$.

The mean difference between the fitnesses of a mutant and a resident is then

$$\begin{aligned}\delta\overline{W}(h, H_{\mathrm{s}}) &= W(h, H_{\mathrm{s}}) - \overline{W}_{\mathrm{r}}(h) \\ &= \left[f\big(\eta(h, H_{\mathrm{s}})\big) - C(h)\right] - \frac{\epsilon(n-1)}{1-\epsilon}\left[f\big(\eta(h, H_{\mathrm{s}})\big) - C(H_{\mathrm{s}})\right] \\ &\quad - \frac{1-n\epsilon}{1-\epsilon}\left[f(\eta_{\mathrm{s}}) - C(H_{\mathrm{s}})\right] \\ &= \frac{1-n\epsilon}{1-\epsilon}\Big(f\big(\eta(h, H_{\mathrm{s}})\big) - f(\eta_{\mathrm{s}})\Big) + C(H_{\mathrm{s}}) - C(h)\,. \end{aligned} \tag{4.17}$$

Noting that $\delta\overline{W}(h, H_{\mathrm{s}}) \to 0$ as $h \to H_{\mathrm{s}}$,

$$\begin{aligned}\frac{d}{dh}\delta\overline{W}(h, H_{\mathrm{s}})|_{h=H_{\mathrm{s}}} &= \lim_{h\to H_{\mathrm{s}}} \frac{\delta\overline{W}(h, H_{\mathrm{s}}) - 0}{h - H_{\mathrm{s}}} \\ &= \lim_{h\to H_{\mathrm{s}}} \frac{1-n\epsilon}{1-\epsilon}\frac{f(\eta(h, H_{\mathrm{s}})) - f(\eta_{\mathrm{s}})}{\eta(h, H_{\mathrm{s}}) - \eta_{\mathrm{s}}} - \frac{C(H_{\mathrm{s}}) - C(h)}{H_{\mathrm{s}} - h} \\ &= \frac{1-n\epsilon}{1-\epsilon} f'(\eta_{\mathrm{s}}) - C'(H_{\mathrm{s}}) \\ &= \frac{1-n\epsilon}{1-\epsilon}\big(f'(\eta_{\mathrm{s}}) - C'(H_{\mathrm{s}})\big) - \frac{(n-1)\epsilon}{1-\epsilon}C'(H_{\mathrm{s}})\,. \end{aligned} \tag{4.18}$$

Because H_{s} is an evolutionarily singular strategy, $f'(nH_{\mathrm{s}}) - C'(H_{\mathrm{s}}) = 0$. If $C'(H_{\mathrm{s}}) > 0$, then

$$\frac{d}{dh}\delta\overline{W}(h, H_{\mathrm{s}})|_{h=H_{\mathrm{s}}} = -\frac{(n-1)\epsilon}{1-\epsilon}C'(H_{\mathrm{s}}) < 0\,. \tag{4.19}$$

It follows that $\delta\overline{W}(h, H_{\mathrm{s}})$ decreases to 0 with h in a left-neighbourhood of $h = H_{\mathrm{s}}$. Hence, mutants playing $h < H_{\mathrm{s}}$ have a fitness advantage over the resident strategy. Similarly, if $C'(H_{\mathrm{s}}) < 0$, mutants playing $h > H_{\mathrm{s}}$ have a fitness advantage over the resident strategy. Consequently, if $C'(H_{\mathrm{s}}) \neq 0$, then H_{s} is not an evolutionarily stable strategy ($\mathrm{ESS}_{\mathrm{N}}$).

The population size does not appear in equation (4.17). Thus, the argument above also shows that in an infinite population containing a proportion ϵ of mutants, when groups contain at most one mutant, then for h close enough to H_{s}, the cumulative effect of the mutants on the residents' fitness causes the mean resident fitness to be smaller than the mutant fitness. Consequently, in this case, H_{s} is also not ES. □

4.6.2 Proof of theorem 4.3.2

We will use the following lemma, proved in appendix 4.C:

Lemma 4.6.1 (Directional selection)**.** *Suppose that the hypotheses of theorem 4.3.2 hold, and that a population of residents playing H is invaded by K mutants playing h.*

- *If $\partial_h \delta\overline{W}(h,H)|_{h=H} = \frac{N-n}{N-1} f'\big(nH\big) - C'(H) < 0$ then less cooperative mutants playing $h < H$ sufficiently close to H will obtain a higher fitness ($\delta\overline{W}(h,H) > 0$) regardless of how many mutants $1 \leq K \leq N-1$ are in the population.*
- *If $\partial_h \delta\overline{W}(h,H)|_{h=H} = \frac{N-n}{N-1} f'\big(nH\big) - C'(H) > 0$ then more cooperative mutants playing $h > H$ sufficiently close to H will obtain a higher fitness ($\delta\overline{W}(h,H) > 0$) regardless of how many mutants $1 \leq K \leq N-1$ are in the population.*

Now recall that if $H_{\mathrm{s}} = \eta_{\mathrm{s}}/n$ is an evolutionarily singular strategy in the infinite-population game, the fitness gradient $D(H) = f'(nH) - C'(H)$ vanishes at H_{s}, so $f'(\eta_{\mathrm{s}}) = f'(nH_{\mathrm{s}}) = C'(H_{\mathrm{s}})$, and consequently,

$$\partial_h \delta\overline{W}(h,H_{\mathrm{s}})|_{h=H_{\mathrm{s}}} = \frac{N-n}{N-1} f'\big(nH_{\mathrm{s}}\big) - C'(H_{\mathrm{s}}) = -\frac{n-1}{N-1} C'(H_{\mathrm{s}})\,. \tag{4.20}$$

Thus, if $C'(H_{\mathrm{s}}) > 0$, then $\frac{N-n}{N-1} f'\big(nH_{\mathrm{s}}\big) - C'(H_{\mathrm{s}}) < 0$, and it follows from lemma 4.6.1 that mutants playing $h < H_{\mathrm{s}}$ sufficiently close to H_{s} obtain a higher fitness than residents and can thus invade. Similarly, if $C'(H_{\mathrm{s}}) < 0$ then mutants playing $h > H_{\mathrm{s}}$ sufficiently close to H_{s} can invade.

Interestingly, irrespective of the proportion of mutants in the population, K/N, there are always levels of contribution h sufficiently close to H_{s} which yield a higher payoff than the resident strategy H_{s}. □

4.6.3 Proof of lemma 4.3.7

Consider a population of residents playing H invaded by one mutant playing h. There is then one group which contains $n-1$ residents and a single mutant (in which the total good contributed is $\eta(h,H) = h + (n-1)H$), and there are $G - 1 = N/n - 1$ groups containing n residents (in which the total good contributed is $\eta(H,H) = nH$). Thus, the mean fitness of agents playing the resident strategy is

$$\overline{W}_{\mathrm{r}}(h,H) = \frac{n-1}{N-1} W(H, \frac{h+(n-2)H}{n-1}) + \frac{N-n}{N-1} W(H,H)\,, \tag{4.21}$$

and the mutant's fitness is $W(h,H)$.

The mean difference between the fitness of a mutant and that of a resident is then

$$\begin{aligned}\delta\overline{W}(h,H) &= W(h,H) - \overline{W}_{\mathrm{r}}(h,H)\\ &= \big[f\big(\eta(h,H)\big) - C(h)\big] - \frac{n-1}{N-1}\big[f\big(\eta(h,H)\big) - C(H)\big] - \frac{N-n}{N-1}\left[f(nH) - C(H)\right]\\ &= \frac{N-n}{N-1}\Big(f\big(h+(n-1)H\big) - f(nH)\Big) + C(H) - C(h)\,. \end{aligned} \tag{4.22}$$

It follows that $\delta\overline{W}(H,H) = 0$, and a mutant is selected against if $\delta\overline{W}(h,H) < 0$. Thus, mutations near $\hat{H} > 0$ are selected against if $h = \hat{H}$ is a strict local maximum of $\delta\overline{W}(h,\hat{H})$ (for fixed $H = \hat{H}$). □

4.6.4 Proof of theorem 4.3.9

If $\delta\overline{W}(h,H)$ is differentiable in h and $\partial_h\delta\overline{W}(h,\hat{H})|_{h=\hat{H}} = 0$ and $\partial_h^2\delta\overline{W}(h,\hat{H})|_{h=\hat{H}} < 0$, then $\delta\overline{W}(h,\hat{H})$ attains a local maximum at $h = \hat{H}$. Thus, from lemma 4.3.7, we have the following condition for selection opposing invasion near $H > 0$:

$$\partial_h\delta\overline{W}(h,\hat{H})|_{h=\hat{H}} = \frac{N-n}{N-1}f'\big(n\hat{H}\big) - C'(\hat{H}) = 0\,, \tag{4.23a}$$

$$\partial_h^2\delta\overline{W}(h,\hat{H})|_{h=\hat{H}} = \frac{N-n}{N-1}f''\big(n\hat{H}\big) - C''(\hat{H}) < 0\,. \tag{4.23b}$$

The above conditions for selection opposing invasion must be modified when the resident strategy is $H = 0$: Selection opposes invasion of $H = 0$ by sufficiently similar strategies if $\delta\overline{W}(h,H) < 0$ only for all $h > 0$ sufficiently close to $h = 0$ (because mutants $h < H = 0$ are biologically meaningless). This is satisfied if either condition (4.23) holds for $\hat{H} = 0$, or if

$$\partial_h\delta\overline{W}(h,0)|_{h=0} = \frac{N-n}{N-1}f'\big(0\big) - C'(0) < 0\,. \tag{4.24}$$

To compare condition (4.23) with the conditions derived from adaptive dynamics, note that as $N \to \infty$, equation (4.23a) approaches $f'(n\hat{H}) - C'(\hat{H}) = 0$, in which case $\hat{H}$ is an evolutionarily signular point (that is, $D(\hat{H}) = f'(n\hat{H}) - C'(H) = 0$). Similarly, as $N \to \infty$, equation (4.23b) approaches $f''\big(n\hat{H}\big) - C''(\hat{H}) < 0$, which is the adaptive dynamics condition for selection opposing invasion at a singular strategy (and thus for evolutionary stability).

We now examine the convergent stability of a strategy $\hat{H} > 0$. If

$$\partial_h\delta\overline{W}(h,H)|_{h=H} > 0\,, \tag{4.25}$$

then mutants contributing slightly more than the residents ($h > H$) obtain a higher fitness, whereas if $\partial_h W(h,H)|_{h=H} < 0$, mutants contributing slightly less than the residents ($h < H$) obtain a higher fitness than the residents. Thus, if $\hat{H}$, $\partial_h \delta\overline{W}(h,H)|_{h=H} > 0$ for $H < \hat{H}$ and $\partial_h \delta\overline{W}(h,H)|_{h=H} < 0$ for $H > \hat{H}$ when H is sufficiently close to $\hat{H}$, then $\hat{H}$ is locally convergently stable. Hence, a sufficient condition for local convergent stability is

$$\partial_H \left[\partial_h \delta\overline{W}(h,H)|_{h=H}\right]_{H=\hat{H}} < 0\,, \tag{4.26}$$

or equivalently

$$\left[\partial_h^2 \delta\overline{W}(h,H) + \partial_h \partial_H \delta\overline{W}(h,H)\right]_{h=H=\hat{H}} < 0\,. \tag{4.27}$$

Using equation (4.22), condition (4.27) becomes

$$n\frac{N-n}{N-1} f''\big(n\hat{H}\big) - C''(\hat{H}) < 0\,. \tag{4.28}$$

Note that equation (4.28) implies that when H is sufficiently close to $\hat{H}$, $\partial_h \delta\overline{W}(h,H)|_{h=H} > 0$ for $H < \hat{H}$ and $\partial_h \delta\overline{W}(h,H)|_{h=H} < 0$ for $H > \hat{H}$. Thus, by lemma 4.6.1, if a mutant strategy h between H and $\hat{H}$ is sufficiently close to H, then the mutants' fitness is larger than the residents', regardless of the number of mutants in the population $1 \leq K \leq N-1$, which implies that selection favours fixation of the mutant strategy (by corollary 5.6.5).

$\hat{H} = 0$ is convergently stable if for sufficiently small $H > 0 = \hat{H}$, mutants contributing slightly less than the resident, $h < H$ obtain a higher fitness than the residents. Thus, if $\partial_h W(h,H)|_{h=H} < 0$ for sufficiently small $H > 0$, then $\hat{H} = 0$ is convergently stable. Using equation (4.22), we conclude that $\hat{H} = 0$ is convergently stable if for sufficiently small $H > 0$,

$$\frac{N-n}{N-1} f'\big(nH\big) - C'(H) < 0\,. \tag{4.29}$$

As for the case $\hat{H} > 0$, this implies that for h sufficiently close to H, selection favours fixation of h. □

Acknowledgments

We were supported by NSERC (DE) and the Ontario Trillium Foundation (CM). We are grateful to Sigal Balshine, Ben Bolker, Michael Doebeli, Jonathan Dushoff, Paul Higgs and Rufus Johnstone for valuable discussions and comments.

Appendix

4.A Analysis frameworks

Two frameworks commonly used in analyzing evolutionary games with continuous strategy sets in infinite populations are static evolutionary game theory[28, 59, 152, 156, 158] and adaptive dynamics [56, 57, 58, 59]. Below, we recall some of the main concepts from these frameworks, as they apply to our analysis. For a general treatment, see the references cited above. We conclude by addressing considerations arising when analyzing games in finite populations.

4.A.1 Static evolutionary game theory in infinite populations

Definition 4.A.1 (Evolutionary stability)**.** *A contribution level $\hat{H} \geq 0$ is **evolutionarily stable** (ES)* iff *a single agent that plays a different strategy cannot invade the population (all strategies different from $\hat{H}$ are selected against).*

As different levels of contributions constitute strategies in this game, we also use the term ***evolutionarily stable strategy*** (ESS), when referring to a level of contribution that is ES.

Since evolution by natural selection typically involves mutations that have a small phenotypic effect, the following definition is also biologically relevant:

Definition 4.A.2 (Local Evolutionary stability)**.** *A contribution level $\hat{H} \geq 0$ is **locally evolutionarily stable** (locally ES) if a single agent playing a mutant strategy h different from, but sufficiently close to $\hat{H}$ cannot invade the population (h is selected against if $|\hat{H}-h|$ is sufficiently small) [59, 157].*

Definition 4.A.3 (Local convergent stability)**.** *A contribution level $\hat{H} \geq 0$ is **locally convergently stable** (locally CS) if, when the resident strategy H is close enough to $\hat{H}$, a mutant playing a strategy between H and $\hat{H}$ and sufficiently close to H can invade the population (h is selected for if $H < h \leq \hat{H}$ or $\hat{H} < h < H$ and h is sufficiently close to H).*

Property	Characterization
Local evolutionary stability	$\frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} < 0$
Convergence stability	$\frac{\partial^2 s_H(h)}{\partial H^2} - \frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} > 0$
Singular strategy can spread in populations playing sufficiently similar strategy	$\frac{\partial^2 s_H(h)}{\partial H^2}\Big\vert_{h=H} > 0$
Mutually-invasible strategies exist near singular point	$\frac{\partial^2 s_H(h)}{\partial H^2} + \frac{\partial^2 s_H(h)}{\partial h^2}\Big\vert_{h=H} > 0$

Table 4.A.1: Local properties of singular strategies in adaptive dynamics, as in [57, Table 1].

4.A.2 Adaptive dynamics

Adaptive dynamics [56, 57, 58] can also be used to gain insight into the evolution of continuous traits in an infinite population. In particular, Doebeli *et al.* [55] use the adaptive dynamics framework to completely characterize the evolutionary dynamics of the continuous snowdrift game with smooth payoffs. Here, we briefly outline concepts from adaptive dynamics.

Following [55, 57], the ***growth rate*** of a rare mutant strategy h in a resident population playing H is

$$s_H(h) = W(h, H) - W(H, H)\,, \tag{4.30}$$

where $W(x, y)$ is the fitness of a mutant playing x in a population playing y. The ***local fitness gradient*** is then

$$D(H) = \frac{\partial s_H(h)}{\partial h}\bigg|_{h=H}\,, \tag{4.31}$$

and the ***adaptive dynamics*** of H are given by

$$\dot{H} = D(H)\,. \tag{4.32}$$

An equilibrium of equation (4.32), that is, $\hat{H}$ satisfying $D(\hat{H}) = 0$, is called a ***singular strategy***. A singular strategy that is an attractor of equation (4.32) is convergently stable in the sense of definition 4.A.3. A singular strategy H can also be locally evolutionarily stable as in definition 4.A.2. The mathematical conditions for these and other possible characteristics of singular strategies are listed in table 4.A.1, following [57].

4.A.3 Finite populations

Genetic drift [61] is a significant force that shapes evolution in finite populations. In an asexual population of constant size N, a neutral mutation fixes (that is, takes over the population) with probability $1/N$ (the probabilities of any one agent being the ancestor of the entire population at some future point in time are identical). Furthermore, the fixation probability of a mutation that is selected against when rare can be larger than $1/N$, if it is selected for when common enough [186]. Motivated by this, Nowak *et al.* [186] have refined the definition of evolutionary stability of a strategy in a finite population so that selection opposes both invasion and fixation of mutant strategies. We apply their approach in the context of continuous strategy sets to obtain the following refinement of the definitions of evolutionary stability given in appendix 4.A.1:

Definition 4.A.4 (Evolutionary stability in a finite population)**.** *A strategy $\hat{H}$ is **locally evolutionarily stable** (local ESS$_N$) in a population of size N* iff*, when a single mutant playing $h \neq \hat{H}$ sufficiently close to $\hat{H}$ arises in a population playing $\hat{H}$,*

- *the mutant's fitness is lower than the residents' (selection opposes invasion),*
- *the mutant's fixation probability is less than $1/N$ (selection opposes fixation).*

If the above holds for any *mutant strategy $h \neq \hat{H}$, then $\hat{H}$ is said to be **globally evolutionarily stable** (global ESS$_N$).*

Remark 4.A.5. *In the adaptive dynamics framework, if a homogeneous population playing H near a singular strategy $\hat{H}$ can be invaded by a mutant between H and $\hat{H}$, then the mutant will fix [201, Proposition 1]. However, in principle, when the scenario just described is played out in a finite population, selection favouring invasion of h does not necessarily imply that selection favours its fixation. Thus, the definition of convergent stability (definition 4.A.3) might also need to be modified for finite populations (to our knowledge, this issue has not been rigorously addressed in the literature). However, in the n-player snowdrift game, this potential issue is a moot point: lemma 4.6.1 implies that if a finite population of agents playing a non-singular strategy H can be invaded by a single mutant playing h, then if h is sufficiently close to H, mutants will have a higher fitness than residents for any number of mutants $1 \leq K \leq N-1$, so by corollary 5.6.5 selection must also favour fixation of h. Thus, we will not have occasion to seek a redefinition of convergent stability for finite populations here.*

4.B Application of finite population theorems to subclass of snowdrift games

Theorems 4.3.1 and 4.3.2 show that the evolutionary dynamics of snowdrift games in finite populations differ from those predicted by infinite-population models. In this appendix, we evaluate the extent of these differences by finding finite-population evolutionarily stable strategies (ESS_N) for a sub-class of snowdrift games, the assumptions of which are outlined in §4.4. When a game from this class is played in an infinite population, there is always a cooperative ESS (theorem 3.5.1), which we denote by H^*_∞ .

To use theorem 4.3.9, we also assume f is twice-differentiable in a neighbourhood of $\eta_{\max}$. This, in conjunction with assumption (c) in §4.4 implies $f''(\eta_{\max}) < 0$. Also, observe that if $\eta_{\min} = 0$, then $f(\eta_{\max}) \geq \eta_{\max}$, so our assumption $f(\eta_{\max}) < \eta_{\max}$ implies $\eta_{\min} > 0$.

First, because $f(\eta) - \eta$ decreases for $0 \leq \eta < \eta_{\min}$, it follows from condition (4.24) that selection opposes invasion of $H = 0$ (defection) by sufficiently similar strategies, and since condition (4.9) is satisfied for sufficiently small $H > 0$, $H = 0$ is locally convergently stable.

In fact, if there are $1 \leq K \leq N - 1$ mutants playing h and $N - K$ mutants playing $H = 0$, then using equations (4.61) and (4.63), we have

$$\partial_h \delta\overline{W}(h, 0)|_{h=0} = -C'(H_\text{s}) + \frac{N-n}{N-1} f'(0) = \frac{N-n}{N-1} f'(0) - 1 \,. \tag{4.33}$$

Because $\eta_{\min} > 0$, we have $f'(0) < 1 < \frac{N-1}{N-n} = 1 + \frac{n-1}{N-n}$, and thus $\partial_h \delta\overline{W}(h, 0)|_{h=0} < 0$, so mutant strategies sufficiently close to $H = 0$ are selected against (for any number of mutants $1 \leq K \leq N - 1$). Thus, corollary 5.6.5 implies that $H = 0$ is a local ESS_N.

Next, observe that for this model, the condition for $H > 0$ being a singular strategy, equation (4.23a), simplifies into

$$f'(nH) = \frac{N-1}{N-n} = 1 + \frac{n-1}{N-n} > 1 \,. \tag{4.34}$$

Because $f'(nH) > 1$ only in the range $\eta_{\min}/n < H < \eta_{\max}/n$, if a cooperative ESS_N exists, it can only occur in this range. Note that in an infinite population, $H^*_\infty = \eta_{\max}/n$ is the unique cooperative ESS, so if there is an ESS in the finite-population game, the infinite-population analysis *overestimates* the ESS_N contribution in the finite population.

However, for a fixed population size N and group size n, if

$$\max_{\eta_{\min} < \eta < \eta_{\max}} f'(\eta) < 1 + \frac{n-1}{N-n} \,, \tag{4.35}$$

then equation (4.23a) has no solution and the game has no cooperative ESS_N. In such cases, finite population dynamics are qualitatively different from those predicted by infinite-population models (in which a cooperative ESS exists), regardless of the selection process (that is, the population-genetic process which determines how individual fitnesses affect strategy frequencies in the population over time, *e.g.,* the Moran or Wright-Fisher processes).

It is interesting to examine whether or not the qualitative discrepancy between the predictions of finite and infinite population models disappears for large population sizes. To that end, we consider increasing group size N either by increasing the number of groups G (while the group size n either varies or remains constant), or by increasing the group size n while keeping the number of groups G constant.

Rewriteing equation (4.34) as

$$f'(nH) = 1 + \frac{1 - 1/n}{G - 1}\,, \tag{4.36}$$

we obtain the following:

- If the population size N is increased by increasing the number of groups G, then because our assumptions imply $\max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) > 1$, then for large enough population sizes N, there is a cooperative singular strategy at which selection opposes invasion (that is, a candidate for an ESS_N). To see this, observe that as $\eta \to \eta_{\max}^-$, $f'(\eta) \to f'(\eta_{\max}) = 1$ and $f''(\eta_{\max}) < 0$, so for sufficiently small $0 < \delta < \eta_{\max} - \eta_{\min}$, $f'(\eta)$ decreases to $f'(\eta_{\max}) = 1$ and $f''(\eta) < 0$ for $\eta_{\max} - \delta < \eta < \eta_{\max}$. Thus, for large enough N (that is, when the number of groups G is large enough), because $1 + \frac{1-1/n}{G-1} \to 1$ as $G \to \infty$, there is a solution to

  $$f(\eta) = 1 + \frac{1 - 1/n}{G - 1}\,, \tag{4.37}$$

  in $(\eta_{\max} - \delta, \eta_{\max})$, which we denote by η_N. Thus,

  $$H_N^* = \eta_N/n \tag{4.38}$$

  solves equation (4.34) and hence is a singular strategy. Because $f''(\eta) < 0$ in $(\eta_{\max} - \delta, \eta_{\max})$. Thus $f''\big(nH_N^*\big) < 0$ and $C''(H) = 0$, so equations (4.67) are satisfied at H_N^*, and theorem 4.D.1 implies that H_N^* is a local ESS_N and is convergently stable.

  However, because $N > n$ and $\frac{n-1}{N-n}$ decreases with N, the largest possible value for the right hand side of equation (4.34) is $1 + \frac{2n-1}{n} = 3 - 1/n \geq 2$, achieved when $N = 2n$ (recall that n divides N). Thus, if

  $$\max_{\eta_{\min}<\eta<\eta_{\max}} f'(\eta) < 2\,, \tag{4.39}$$

then for sufficiently small population size (in particular, $N = 2n$) there is no cooperative singular strategy, and thus no ESS_N.

- Now suppose that the population size N and group size n are increased while the number of groups $G = N/n$ remains constant. If

$$\max_{\eta_\text{min}<\eta<\eta_\text{max}} f'(\eta) < 1 + \frac{1-1/n}{G-1}\,, \tag{4.40}$$

for some $n_0 > 1$, then because $1 + \frac{1-1/n}{G-1}$ increases with n, equation (4.36) has no solution for any $n > n_0$. Thus, if group size is increased along with population size while the number of groups is constant, it is possible (if condition (4.40) is satisfied) that for arbitrarily large N, there will be no cooperative singular strategy, and thus no ESS_N.

In fact, if

$$1 + \frac{1}{2(G-1)} = 1 + \frac{1-1/n}{G-1}\bigg|_{n=2} < \max_{\eta_\text{min}<\eta<\eta_\text{max}} f'(\eta) < 1 + \frac{1}{G-1}\,, \tag{4.41}$$

then when the number of groups G is fixed, for low group sizes there is a cooperative ESS singular strategy, but as group size n (and consequently population size N) is increased, there is a group size beyond which $\max_{\eta_\text{min}<\eta<\eta_\text{max}} f'(\eta) < \frac{1-1/n}{G-1}$, and so there is no cooperative ESS_N (even for arbitrarily large group and population sizes). If, for a fixed number of groups G,

$$\max_{\eta_\text{min}<\eta<\eta_\text{max}} f'(\eta) < 1 + \frac{1}{2(G-1)}\,, \tag{4.42}$$

there is no cooperative ESS_N for any population size.

To determine the evolutionary outcome when there is no cooperative ESS_N, that is, when $f'(\eta) < \frac{N-1}{N-n}$ for all $\eta_\text{min} < \eta < \eta_\text{max}$, recall first that $H = 0$ is a local ESS_N. We now substitute $C(h) = h$ in equation (4.22) and differentiate to obtain

$$\delta\overline{W}(h,H) = \frac{N-n}{N-1}\Big(f\big(h+(n-1)H\big) - f(nH)\Big) + H - h \tag{4.43a}$$

$$\partial_h\delta\overline{W}(h,H) = \frac{N-n}{N-1}f'\big(h+(n-1)H\big) - 1\,. \tag{4.43b}$$

Because $f(\eta) - \eta$ decreases for $0 < \eta < \eta_\text{min}$ and $\eta > \eta_\text{max}$, $f'(\eta) \leq 1$ for $0 < \eta \leq \eta_\text{min}$ and $\eta \geq \eta_\text{max}$. It now follows from equations (4.43) that $\delta\overline{W}(h,H)$ decreases with h for any H. Thus defection ($H = 0$) is a globally convergently stable ESS_N.

4.C Proof of lemma 4.6.1

We begin by calculating the proportion of mutants in the population that are in a group containing k mutants. Choose an agent at random from the population by first choosing a group at random and then choosing an agent at random from within that group. Let I be an indicator for whether the chosen agent is a mutant ($I = 1$ if the chosen agent is a mutant, and $I = 0$ otherwise). Let M be the number of mutants in the chosen group. We use Bayes' Theorem [167] to find $\Pr(M = k|I = 1)$, that is, the probability that a chosen mutant is in a group containing k mutants:

$$\Pr(M = k|I = 1) = \frac{\Pr(M = k)\Pr(I = 1|M = k)}{\Pr(I = 1)}\,, \tag{4.44a}$$

$$\Pr(M = k|I = 0) = \frac{\Pr(M = k)\Pr(I = 0|M = k)}{\Pr(I = 0)}\,. \tag{4.44b}$$

If a group of n agents is randomly drawn (without replacement) from a population of N agents consisting of K mutants and $N - K$ residents, the number of mutants in the sampled group is hypergeometrically distributed with parameters N, K and n [167]. That is, the probability of k mutants occurring in a random sample of n agents is:

$$\Pr(M = k) = \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N}{n}}\,. \tag{4.45}$$

Using equation (4.44), we have

$$\Pr(M = k|I = 1) = \frac{k/n}{K/N}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N}{n}} = \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}}\,, \tag{4.46}$$

$$\Pr(M = k|I = 0) = \frac{(n-k)/n}{(N-K)/N}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N}{n}} = \frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}}\,. \tag{4.47}$$

If mutants play h and residents play H, the payoffs to mutants and residents in groups containing k mutants are (respectively)

$$W_{\mathrm{m},k}(h) = W\left(h, \frac{(k-1)h + (n-k)H}{n-1}\right) = f\big(kh + (n-k)H\big) - C(h)\,, \tag{4.48}$$

and

$$W_{\mathrm{r},k}(h) = W\left(H, \frac{kh + (n-1-k)H}{n-1}\right) = f\big(kh + (n-k)H\big) - C(H)\,. \tag{4.49}$$

The mean mutant fitness is then

$$\overline{W}_{\mathrm{m}}(h) = \sum_{k=1}^{n} \Pr\big(M = k | I = 1\big) W_{\mathrm{m},k}(h) = \sum_{k=1}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} W_{\mathrm{m},k}(h)\,, \tag{4.50}$$

and the mean resident fitness is

$$\overline{W}_{\mathrm{r}}(h) = \sum_{k=0}^{n-1} \Pr\big(M = k | I = 0\big) W_{\mathrm{r},k}(h) = \sum_{k=0}^{n-1} \frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}} W_{\mathrm{r},k}(h)\,. \tag{4.51}$$

Note that

$$\sum_{k=1}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} = \sum_{k=0}^{n-1} \frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} = 1 \tag{4.52}$$

(because the sum of the probabilities of all the possible number of mutants occurring in a group of size $n-1$ is one). So using

$$\binom{K-1}{k-1} = \frac{(K-1)!}{(k-1)!(K-k)!} = \frac{k}{K}\frac{K!}{k!(K-k)!} = \frac{k}{K}\binom{K}{k}\,, \tag{4.53}$$

as well as equations (4.52) and (4.48), the mean mutant payoff (equation (4.50)) can also be written as:

$$\begin{aligned}
\overline{W}_{\mathrm{m}}(h) &= \sum_{k=1}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} W_{\mathrm{m},k}(h) \\
&= \sum_{k=1}^{n} \left\{ \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} f\big(kh + (n-k)H\big) - C(h) \right\} \\
&= \sum_{k=1}^{n} \left\{ \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} f\big(kh + (n-k)H\big) \right\} - C(h) \\
&= \sum_{k=1}^{n} \left\{ \frac{k}{K}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} f\big(kh + (n-k)H\big) \right\} - C(h)\,.
\end{aligned} \tag{4.54}$$

Similarly, using

$$\sum_{k=0}^{n-1} \frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}} = 1 \tag{4.55}$$

and

$$\binom{N-K-1}{n-k-1} = \frac{(N-K-1)!}{(n-k-1)!\big(N-K-(n-k)\big)!} = \frac{n-k}{N-K} \times \frac{(N-K)!}{(n-k)!\big(N-K-(n-k)\big)!} = \frac{n-k}{N-K}\binom{N-K}{n-k}, \tag{4.56}$$

equation (4.49) becomes

$$\begin{aligned}\overline{W}_{\mathrm{r}}(h) &= \sum_{k=0}^{n-1}\left\{\frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}} f\big(kh+(n-1)H\big)\right\} - C(H) \\ &= \sum_{k=0}^{n-1}\left\{\frac{n-k}{N-K}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} f\big(kh+(n-1)H\big)\right\} - C(H)\,. \end{aligned} \tag{4.57}$$

Thus, the mean difference between the mutant and resident fitnesses is

$$\begin{aligned}\delta\overline{W}(h,H) &= \frac{\binom{K-1}{n-1}}{\binom{N-1}{n-1}} f(nh) - \frac{\binom{N-K-1}{n-1}}{\binom{N-1}{n-1}} f(\eta) + C(H) - C(h) \\ &\quad + \sum_{k=1}^{n-1} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \left(\frac{k}{K} - \frac{n-k}{N-K}\right) f\big(kh+(n-k)H\big) \\ &= \frac{\binom{K-1}{n-1}}{\binom{N-1}{n-1}} f(nh) - \frac{\binom{N-K-1}{n-1}}{\binom{N-1}{n-1}} f(\eta) + C(H) - C(h) \\ &\quad + \sum_{k=1}^{n-1} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} f\big(kh+(n-k)H\big)\,. \end{aligned} \tag{4.58}$$

Noting that

$$\frac{\binom{K-1}{n-1}}{\binom{N-1}{n-1}} f(nh) = \frac{\binom{N-K}{n-n}\binom{K}{n}}{\binom{N-1}{n-1}} \frac{nN-Kn}{K(N-K)} f\big(nh+(n-n)H\big)\,, \tag{4.59}$$

$$-\frac{\binom{N-K-1}{n-1}}{\binom{N-1}{n-1}} f(\eta) = \frac{\binom{N-K}{n-0}\binom{K}{0}}{\binom{N-1}{n-1}} \frac{0N-Kn}{K(N-K)} f\big(0h+(n-0)H\big)\,, \tag{4.60}$$

we have

$$\delta\overline{W}(h,H) = C(H) - C(h) + \sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} f\big(kh+(n-k)H\big)\,, \tag{4.61}$$

(note that when $K = 1$, equations (4.17) and (4.61) are identical). Thus,

$$\partial_h \delta\overline{W}(h,H)|_{h=H} = -C'(H) + f'(nH)\sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} k\,, \tag{4.62}$$

It can be shown (see appendix 4.F) that

$$\sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} \frac{kN-Kn}{(N-K)} = \frac{N-n}{N-1}\,, \tag{4.63}$$

so,

$$\partial_h \delta\overline{W}(h,H)|_{h=H} = \frac{N-n}{N-1} f'(H) - C'(H)\,. \tag{4.64}$$

Now, if

$$\frac{N-n}{N-1} f'\big(n\hat{H}\big) - C'(\hat{H}) < 0\,, \tag{4.65}$$

then for any population size $N > 1$ and group size $n > 1$, $\partial_h \delta\overline{W}(h,H)|_{h=H} < 0$, regardless of the number of mutants in the population, $1 \leq K \leq N-1$. Because $\delta\overline{W}(H,H) = 0$, it follows that mutants playing $h < H$ sufficiently close to H obtain a higher fitness than residents.

Similarly, if

$$\frac{N-n}{N-1} f'\big(n\hat{H}\big) - C'(\hat{H}) > 0\,, \tag{4.66}$$

then for any group size $n > 1$ and population size $N > 1$ and for any number of mutant $1 \leq K \leq N-1$ in the population, $\partial_h \delta\overline{W}(h,H)|_{h=H} > 0$. Because $\delta\overline{W}(H,H) = 0$, it follows that mutants playing $h > H$ sufficiently close to H obtain a higher fitness than residents, which completes the proof.

4.D Sufficient condition for evolutionary stability in the continuous snowdrift game in a finite population

Theorem 4.D.1. *Suppose that a population of size N is playing the public goods game described in theorem 4.3.1.*

A singular strategy $H > 0$ satisfying

$$\frac{N-n}{N-1} f''\big(nH\big) < C''(H)\,, \tag{4.67a}$$

and

$$(2n-1)\frac{N-n}{N-1}f''\big(nH\big) < C''(H)\,, \tag{4.67b}$$

is a convergently stable local ESS$_N$, regardless of the selection process at work.

Proof. Consider the invasion of a population playing the singular strategy H by mutants playing a strategy $h \neq H$.

Selection opposes invasion of H sufficiently close to H if equation (4.23b) holds, that is,

$$\frac{N-n}{N-1}f''\big(nH\big) - C''(H) < 0\,. \tag{4.68}$$

We now wish to ensure that selection opposes fixation of mutants playing $h \neq H$ sufficiently close to H. Differentiating equation (4.61) again with respect to h, we have

$$\partial_h^2 \delta\overline{W}(h,H)|_{h=H} = -C''(H) + \sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} k^2 f''\big(nH\big)\,. \tag{4.69}$$

Recalling that

$$\sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} k^2 = \frac{(N-n)\big(2(K-1)(n-1)+N-2\big)}{(N-1)(N-2)}\,, \tag{4.70}$$

(see appendix 4.F), we have

$$\partial_h^2 \delta\overline{W}(h,H)|_{h=H} = -C''(H) + \left(\frac{2(K-1)(n-1)}{N-2}+1\right)\frac{N-n}{N-1}f''\big(nH\big)\,, \tag{4.71}$$

so $\partial_h^2 \delta\overline{W}(h,H)|_{h=H} < 0$ *iff*

$$\left(\frac{2(K-1)(n-1)}{N-2}+1\right)\frac{N-n}{N-1}f''\big(nH\big) < C''(H)\,. \tag{4.72}$$

For a fixed number of mutants $1 \leq K \leq N-1$, if $\partial_h^2 \delta\overline{W}(h,H)|_{h=H} < 0$, then $h = H$ is a local maximum of $\delta\overline{W}(h,H)$ so mutants playing h sufficiently close to H are selected against. Thus, if equation (4.71), holds for any number of mutants $1 \leq K \leq N-1$, mutants are selected against, regardless of their frequency in the population.

If $N > 2$ and $1 \leq K < N$, the coefficient $\frac{2(K-1)(n-1)}{N-2}+1$ increases with K from 1 to $2n-1$. Thus, the right hand side of equation (4.73a) is linear in K, and decreases if $f''(nH) < 0$, increases if $f''(nH) > 0$, and is constant if $f''(nH) = 0$. Consequently,

equation (4.73a) holds for all $1 \leq K \leq N-1$ *iff* it holds for $K = 1$ and $K = N-1$, that is, *iff*

$$\frac{N-n}{N-1} f''\big(nH\big) < C''(H)\,, \tag{4.73a}$$

and

$$(2n-1)\frac{N-n}{N-1} f''\big(nH\big) < C''(H)\,. \tag{4.73b}$$

Note that equation (4.73a) is identical to the condition for selection opposing invasion by sufficiently similar mutant strategies (equation (4.23b)). Thus, mutants are selected against regardless of their frequency in the population if equations (4.73a) and (4.73b) hold. By corollary 5.6.5, the probability that the mutant fixes is less than $1/N$, so selection also opposes fixation of sufficiently similar mutants. Lastly, if equations (4.73a) and (4.73b) hold, then condition (4.7) is also satisfied, so H is convergently stable. □

4.E Consistency with infinite population limit

The predictions of theorems 4.3.1 and 4.3.2 for a finite population are diametrically opposed to what can be said when the population is infinite: when a continuous n-player snowdrift game from the sub-class defined in chapter 3 is played in an infinite population, theorems 3.5.1 and 3.5.4 predict the existence of a global ESS, $H^*_\infty > 0$, which must be a singular strategy in the infinite population. But theorem 4.3.1 implies that when the same game is played in a finite population, strategies that are singular in an infinite population can be invaded by sufficiently similar strategies, and so H^*_∞ cannot be an ESS_N. Moreover, theorem 3.5.4 shows that in an infinite population, regardless of what proportion ϵ of the population plays the mutant strategy (h), if h is sufficiently close to H^*_∞, then the mutants' mean fitness is lower than the residents'. But when the population is finite, it follows from theorem 4.3.2 that whatever the proportion of mutants in the population ($\epsilon = K/N$), if the mutant strategy h is sufficiently close to H^*_∞, the mean mutant fitness is higher than the mean resident fitness.

To understand this discrepancy, we make the additional assumptions defining the snowdrift games analyzed in chapter 3, and outlined in § 4.4. Recall that when this game is played in an infinite population, these assumptions ensure the existence of a cooperative ESS, $H^*_\infty = \eta_{\max}/n$. To compare with theorems 4.3.1 and 4.3.2, we also assume f is twice-differentiable in a neighbourhood of $\eta_{\max}$. For benefit functions in the class analyzed in chapter 3, this implies $f''(\eta_{\max}) < 0$.

§ 4.E.1 relates the mean fitness difference between a mutant and a resident in the finite and infinite population cases, and §§4.E.2 and 4.E.3 compare the conclusions of theorems 4.3.1 and 4.3.2 with their analogues in an infinite population obtained in chapter 3.

4.E.1 $\delta\overline{W}$ in the infinite population limit

Equation (3.44) gives the difference between the mean mutant and resident fitnesses, when the residents play the (infinite population) ESS H^*_∞. Here, we show that equation (3.44) is obtained as the limit of this same fitness difference defined for a finite population, $\delta\overline{W}(h, H^*_\infty)$ (see equation (4.58)), as the population size $N \to \infty$.

If the population is composed of K mutants playing h and $N-K$ residents playing H^*_∞, using

$$\binom{N-K}{n-1-k} = \frac{(N-K)!}{(n-1-k)!\big(N-K-(n-k)+1\big)!} \tag{4.74}$$

$$= \frac{N-K}{(N-K-(n-k)+1)} \frac{(N-K-1)!}{(n-1-k)!\big(N-K-(n-k)\big)!} \tag{4.75}$$

$$= \frac{N-K}{(N-K-(n-k)+1)} \binom{N-K-1}{n-k-1} \tag{4.76}$$

and

$$\binom{K-1}{k} = \frac{(K-1)!}{k!(K-k-1)!} = \frac{K-k}{K}\binom{K}{k}, \tag{4.77}$$

we can rewrite equation (4.50) as

$$\begin{aligned}
\overline{W}_{\mathrm{m}}(h) &= \sum_{k=1}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} W_{\mathrm{m},k}(h) \\
&= \sum_{k=0}^{n-1} \frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} W_{\mathrm{m},k+1}(h) \\
&= \sum_{k=0}^{n-1} \frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{K-k}{K} \times \frac{N-K}{N-K-(n-k-1)} W_{\mathrm{m},k+1}(h)\,,
\end{aligned} \tag{4.78}$$

where $W_{\mathrm{m},k+1}(h)$ is defined by equation (4.48). Taking the difference of equations (4.78) and (4.51) gives the alternative expression for $\delta\overline{W}(h, H^*_\infty)$,

$$\begin{aligned}
\delta\overline{W}(h, H^*_\infty) &= \overline{W}_{\mathrm{m}}(h) - \overline{W}_{\mathrm{r}}(h) \\
&= \sum_{k=0}^{n-1} \frac{\binom{N-K-1}{n-k-1}\binom{K}{k}}{\binom{N-1}{n-1}} \left[\frac{K-k}{K} \times \frac{N-K}{N-K-(n-k-1)} W_{\mathrm{m},k+1} - W_{\mathrm{r},k} \right].
\end{aligned} \tag{4.79}$$

Recall that in the limit $N \to \infty$, when k, n and $\epsilon = K/N$ are kept fixed, we have

$$\frac{\binom{N-1-K}{n-1-k}\binom{K}{k}}{\binom{N-1}{n-1}} \to \binom{n-1}{k}\epsilon^k(1-\epsilon)^{n-1-k}\,, \tag{4.80}$$

(see [167, p. 161]), and

$$\frac{K-k}{K} \times \frac{N-K}{N-K-(n-k-1)} \to 1 \times 1 = 1 \tag{4.81}$$

Thus, in this limit, equation (4.79) becomes

$$\delta\overline{W}(h, H^*_\infty) = \sum_{k=0}^{n-1}\binom{n-1}{k}\epsilon^k(1-\epsilon)^{n-1-k}\Big[W_{\mathrm{m},k+1} - W_{\mathrm{r},k}\Big]\,, \tag{4.82}$$

that is, equation (3.44) (where the notation $\delta\overline{W}(h)$ was used and $H = H^*_\infty$ was implicit).

4.E.2 Theorem 4.3.1 in the infinite population limit

The applicability of theorem 4.3.1 to an infinite population seems to contradict theorem 3.5.1, in which it is shown that a single agent attempting to invade an infinite population playing the ESS H^*_∞ obtains a lower fitness than the residents. The discrepancy arises because when the population is invaded by a proportion ϵ of mutants, the effect of these mutants on the mean resident fitness is non-negligible. However, because in theorem 4.3.1, each interacting group of n agents contains at most one mutant, the mutant's fitness is unaffected by the presence of other mutants (in contrast to theorem 3.5.4). In the limit of one invader in an infinite population, $\epsilon \to 0$, and

$$\lim_{\epsilon\to 0}\frac{d}{dh}\delta\overline{W}(h, H_{\mathrm{s}})|_{h=H_{\mathrm{s}}} = 0\,, \tag{4.83}$$

so the conclusions of theorem 4.3.1 do not apply.

4.E.3 Theorem 4.3.2 in the infinite population limit

Consider a dimorphic population in which residents play the infinite-population ESS $H = H^*_\infty$, and mutants play h. When the population is infinite, this scenario is analyzed in theorem 3.5.4. When the population is finite, theorem 4.3.2 applies (because H^*_∞ is a singular strategy). Here, we reexamine the proof of theorem 4.3.2 for the singular strategy $H_{\mathrm{s}} = H^*_\infty$. We identify why its conclusions do not hold in the infinite population limit ($N \to \infty$), and thus resolve the apparent contradiction between theorem 4.3.2 and theo-

rem 3.5.4.

Analogously to $\delta\overline{W}(h, H^*_\infty)$, let $\delta\overline{W}_\infty(h, H^*_\infty)$ be the mean difference between the mutant and resident fitnesses in an infinite population (calculated in § 3.8.1). Thus, if the proportion of mutants in the population is ϵ,

$$\delta\overline{W}_\infty(h, H^*_\infty) = \overline{W}_{\mathrm{m}\,\infty}(h) - \overline{W}_{\mathrm{r}\,\infty}(h)\,, \tag{4.84}$$

where

$$\begin{aligned}\overline{W}_{\mathrm{m}\,\infty}(h) &= \sum_{k=0}^{n-1}\binom{n-1}{k}\epsilon^k(1-\epsilon)^{n-1-k}W_{\mathrm{m},k+1}(h)\\ \overline{W}_{\mathrm{r}\,\infty}(h) &= \sum_{k=0}^{n-1}\binom{n-1}{k}\epsilon^k(1-\epsilon)^{n-1-k}W_{\mathrm{r},k}(h)\end{aligned} \tag{4.85}$$

Because H^*_∞ is an ESS in an infinite population, $\delta\overline{W}_\infty(h, H^*_\infty)$ attains a local maximum when $h = H^*_\infty$, so

$$\partial_h\delta\overline{W}_\infty(h, H^*_\infty)|_{h=H^*_\infty} = 0\,. \tag{4.86}$$

We see that this is consistent with taking $N \to \infty$ in equation (4.64):

$$\lim_{N\to\infty}\partial_h\delta\overline{W}(h, H^*_\infty)|_{h=H^*_\infty} = \lim_{N\to\infty}\frac{1-n}{N-1} = 0 = \partial_h\delta\overline{W}_\infty(h, H^*_\infty)|_{h=H^*_\infty}\,, \tag{4.87}$$

so the conclusions of theorem 4.3.2 do not apply in the limit $N \to \infty$.

Differentiating equation (4.61) twice at $h = H^*_\infty$, we get

$$\partial_h^2\delta\overline{W}(h, H^*_\infty)|_{h=H^*_\infty} = f''(\eta_{\max})\sum_{k=0}^{n}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}}\frac{kN-Kn}{K(N-K)}k^2\,. \tag{4.88}$$

In appendix 4.F, we show that

$$\sum_{k=0}^{n}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}}\frac{kN-Kn}{K(N-K)}k^2 = \frac{(N-n)\big(2(K-1)(n-1)+N-2\big)}{(N-1)(N-2)}\,, \tag{4.89}$$

which is positive if $N > n \geq 1$ and $K \geq 1$ (that is, the population consists of more than one group of $n \geq 1$ agents playing the public goods game, and there is at least one mutant in the population). Hence, since $f''(\eta_{\max}) < 0$,

$$\partial_h\delta\overline{W}(h, H^*_\infty)|_{h=H^*_\infty} = f''(\eta_{\max})\frac{(N-n)\big(2(K-1)(n-1)+N-2\big)}{(N-1)(N-2)} < 0\,, \tag{4.90}$$

if $N > 2$ and $1 \leq K < N$.

As the population size goes to $N \to \infty$, while $K/N \to \epsilon$, we have

$$\begin{aligned}\lim_{\substack{N\to\infty \\ K/N\to\epsilon}} \partial_h \delta\overline{W}(h, H^*_\infty)|_{h=H^*_\infty} &= f''(\eta_{\max}) \lim_{\substack{N\to\infty \\ K/N\to\epsilon}} \frac{(1-\frac{n}{N})\big(2(\frac{K}{N}-\frac{1}{N})(n-1)+1-\frac{2}{N}\big)}{(1-\frac{1}{N})(1-\frac{2}{N})} \\ &= f''(\eta_{\max})(1-\epsilon)\big(2\epsilon(n-1)+1\big) < 0\,, \end{aligned} \tag{4.91}$$

which is again consistent with $h = H^*_\infty$ being a maximum of $\delta\overline{W}_\infty(h, H^*_\infty)$.

Because $\partial_h \delta\overline{W}(h, H^*_\infty)|_{h=H^*_\infty} < 0$ for any population size N (see equation (4.64)), it follows that for $h > H^*_\infty$ sufficiently close to H^*_∞, $\partial_h \delta\overline{W}(h, H^*_\infty) < 0$ and consequently $\delta\overline{W}(h, H^*_\infty) < 0 = \delta\overline{W}(H^*_\infty, H^*_\infty)$. However, for $h < H^*_\infty$ sufficiently close to H^*_∞, $\delta\overline{W}(h, H^*_\infty) > \delta\overline{W}(H^*_\infty, H^*_\infty) = 0$ for any population size N, despite the fact that $\delta\overline{W}_\infty(h, H^*_\infty) < \delta\overline{W}_\infty(H^*_\infty, H^*_\infty) = 0$ in the infinite-population limit ($N \to \infty$).

4.F Proofs of equation (4.63) and equation (4.89)

In this appendix, we prove equation (4.63) and equation (4.89).

Let X be hypergeometrically distributed with parameters T, R and m. In the standard interpretation, X describes the number of red balls obtained when m balls are sampled randomly without replacement from an urn containing a total of T balls, R of which are red and $T - R$ are white. Then, the probability of drawing r red balls in a set of m is

$$\Pr(X = r) = \frac{\binom{R}{r}\binom{T-R}{m-r}}{\binom{T}{m}}\,. \tag{4.92}$$

Because the sum of the probabilities of all possible outcomes is 1,

$$\sum_{r=0}^{m} \Pr(X = r) = \sum_{r=0}^{m} \frac{\binom{R}{r}\binom{T-R}{m-r}}{\binom{T}{m}} = 1\,. \tag{4.93}$$

The two first moments of X are then [167, p. 162]

$$\mathbb{E}\,(X) = \sum_{r=0}^{m} r \Pr(X = r) = \sum_{r=0}^{m} r \frac{\binom{R}{r}\binom{T-R}{m-r}}{\binom{T}{m}} = m\frac{R}{T} \tag{4.94}$$

$$\mathbb{E}\,(X^2) = \sum_{r=0}^{m} r^2 \Pr(X = r) = \sum_{r=0}^{m} r^2 \frac{\binom{R}{r}\binom{T-R}{m-r}}{\binom{T}{m}} = m\frac{R}{T}\left(\frac{(m-1)(R-1)}{T-1} + 1\right)\,. \tag{4.95}$$

Observe that

$$\begin{aligned}\sum_{k=0}^{n}\frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}}\frac{kN-Kn}{(N-K)} &= \sum_{k=1}^{n}\frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}}\frac{kN-Kn}{(N-K)}\\ &= \sum_{k=0}^{n-1}\frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}}\frac{(k+1)N-Kn}{(N-K)}\\ &= \sum_{k=0}^{n-1}\frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}}\frac{Nk+N-Kn}{(N-K)}\,,\end{aligned}$$

equations (4.93) and (4.94) with $T = N-1$, $R = K-1$ and $m = n-1$ give

$$\sum_{r=0}^{n-1}\frac{\binom{K-1}{r}\binom{N-K}{n-1-r}}{\binom{N-1}{m}} = 1\,, \tag{4.96}$$

$$\sum_{r=0}^{n-1} r\frac{\binom{K-1}{r}\binom{N-K}{n-1-r}}{\binom{N-1}{n-1}} = (n-1)\frac{K-1}{N-1}\,, \tag{4.97}$$

respectively, so the left-hand side of equation (4.63) turns into

$$\begin{aligned}\sum_{k=0}^{n}\frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}}\frac{kN-Kn}{(N-K)} &= \frac{N}{(N-K)}\sum_{k=0}^{n-1}k\frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} + \frac{N-Kn}{(N-K)}\sum_{k=0}^{n-1}\frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}}\\ &= \frac{N}{(N-K)}(n-1)\frac{K-1}{N-1} + \frac{N-Kn}{(N-K)} &(4.98)\\ &= \frac{N(n-1)(K-1)+(N-Kn)(N-1)}{(N-K)(N-1)} &(4.99)\\ &= \frac{\big(KNn-Nn-NK+N\big)+\big(N^2-N-KNn+Kn\big)}{(N-K)(N-1)} &(4.100)\\ &= \frac{N^2-Nn-NK+Kn}{(N-K)(N-1)} = \frac{(N-K)(N-n)}{(N-K)(N-1)} = \frac{N-n}{N-1}\,, &(4.101)\end{aligned}$$

thus proving equation (4.63).

We now turn to proving equation (4.89), namely:

$$\sum_{k=0}^{n}\frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}}\frac{kN-Kn}{K(N-K)}k^2 = \frac{(N-n)\big(2(K-1)(n-1)+N-2\big)}{(N-1)(N-2)}\,. \tag{4.102}$$

Because $\binom{K}{k} = \binom{K-1}{k-1}\frac{K}{k}$,

$$\begin{aligned}
\sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} k^2 &= \sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} \frac{Nk^2-Knk}{N-K} \\
&= \sum_{k=1}^{n} \frac{\binom{N-K}{n-k}\binom{K-1}{k-1}}{\binom{N-1}{n-1}} \frac{Nk^2-Knk}{N-K} \\
&= \sum_{k=0}^{n-1} \frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} \frac{N(k+1)^2-Kn(k+1)}{N-K} \\
&= \sum_{k=0}^{n-1} \frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} \frac{Nk^2+(2N-Kn)k+N-Kn}{N-K} .
\end{aligned} \tag{4.103}$$

Substituting $T = N-1$, $R = K-1$ and $m = n-1$ into equation (4.95) gives

$$\sum_{r=0}^{n-1} r^2 \frac{\binom{K-1}{r}\binom{N-K}{n-1-r}}{\binom{N-1}{n-1}} = (n-1)\frac{K-1}{N-1}\left(\frac{(n-2)(K-2)}{N-2}+1\right) . \tag{4.104}$$

Using equations (4.96), (4.97) and (4.104), equation (4.103) then becomes

$$\begin{aligned}
\sum_{k=0}^{n} \frac{\binom{N-K}{n-k}\binom{K}{k}}{\binom{N-1}{n-1}} \frac{kN-Kn}{K(N-K)} k^2 &= \sum_{k=0}^{n-1} \frac{\binom{N-K}{n-1-k}\binom{K-1}{k}}{\binom{N-1}{n-1}} \frac{Nk^2+(2N-Kn)k+N-Kn}{N-K} \\
&= \frac{N}{N-K} \times (n-1)\frac{K-1}{N-1}\left(\frac{(n-2)(K-2)}{N-2}+1\right) \\
&\quad + \frac{2N-Kn}{N-K} \times (n-1)\frac{K-1}{N-1} \\
&\quad + \frac{N-Kn}{N-K} \times 1 \\
&= \frac{(N-n)\big(2(K-1)(n-1)+N-2\big)}{(N-1)(N-2)} ,
\end{aligned} \tag{4.105}$$

which completes the proof of equation (4.89).

4.G The mean number of mutants in a mutant's and resident's group in infinite and finite populations

In a population of $N-K$ residents and K mutants, the probability that of the $n-1$ other agents in a randomly chosen agent's group, there are k mutants, given that the chosen agent is a mutant, is given by equation (4.46), or equivalently,

$$\Pr(M = k+1 | I = 1) = \frac{\binom{(N-1)-(K-1)}{(n-1)-k}\binom{K-1}{k}}{\binom{N-1}{n-1}}, \tag{4.106}$$

for $0 \leq k \leq n-1$. It follows that the number of *other* (*i.e.,* nonfocal) mutants in a randomly chosen mutant's group is hypergeometrically distributed and has mean $M_{\text{mut},N} = (n-1)\frac{K-1}{N-1}$ (see equation (4.94)).

Analogously, from equation (4.47), the probability that a randomly chosen resident's group contains k mutants is

$$\Pr(M = k | I = 0) = \frac{\binom{(N-1)-K}{(n-1)-k}\binom{K}{k}}{\binom{N-1}{n-1}}, \tag{4.107}$$

for $0 \leq k \leq n-1$, so the number of mutants in a resident's group is hypergeometrically distributed and has mean $M_{\text{res},N} = (n-1)\frac{K}{N-1}$.

However, in an infinite population in which the proportions of mutants and residents are $\epsilon > 0$ and $1-\epsilon$ (respectively), the probability that k of the $n-1$ agents in a randomly chosen agent's group are mutants, given that the chosen agent is a mutant or a resident, is (respectively)

$$\Pr(M = k+1 | I = 1) = \binom{n-1}{k}\epsilon^k(1-\epsilon)^{(n-1)-k}, \tag{4.108}$$

$$\Pr(M = k | I = 0) = \binom{n-1}{k}\epsilon^k(1-\epsilon)^{(n-1)-k} \tag{4.109}$$

for $0 \leq k \leq n-1$ (see equations (3.41) and (3.42)). Thus, for a randomly chosen focal agent, the number of nonfocal agents in its group who are mutants is binomially distributed with mean $M_{\text{nonfocal},\infty} = (n-1)\epsilon$, regardless of whether the focal agent is a resident or a mutant.

For a given proportion of mutants, $\epsilon = K/N > 0$, if $\epsilon < 1$, then

$$M_{\text{mut},N} = (n-1)\frac{K-1}{N-1} < (n-1)\frac{K}{N} = M_{\text{nonfocal},\infty}, \tag{4.110}$$

and

$$M_{\text{res},N} = (n-1)\frac{K}{N-1} > (n-1)\frac{K}{N} = M_{\text{nonfocal},\infty}\,, \tag{4.111}$$

so on average, mutants interact with fewer mutants, and residents interact with more mutants in a finite population than in an infinite one.

4.H Dependence of the ES contribution on the population size

Let H_N^* be a singular strategy in the game described in theorem 4.3.9. Thus, H_N^* satisfies equation (4.23a). In order to study the dependence of H_N^* on N, we multiply equation (4.23a) by $N-1$ and take the derivative with respect to N, yielding

$$f'\big(nH_N^*\big) - C'(H_N^*) + \big(n(N-n)f''\big(nH_N^*\big) - (N-1)C''(H_N^*)\big)\partial_N H_N^* = 0\,. \tag{4.112}$$

Using equation (4.23a), this is equivalent to

$$\left(n\frac{N-n}{N-1}f''\big(nH_N^*\big) - C''\big(H_N^*\big)\right)\partial_N H_N^* = -\frac{n-1}{(N-1)^2}f'\big(nH_N^*\big)\,. \tag{4.113}$$

The right hand side of equation (4.113) is negative, but the sign of the term

$$n\frac{N-n}{N-1}f''\big(nH_N^*\big) - C''(H_N^*) \tag{4.114}$$

may vary with N and n. Hence, in general, the sign of $\partial_N H_N^*$ is thus unknown. However, if H_N^* satisfies condition (4.7) for convergent stability or any population size N, then, $\partial_N H_N^* > 0$, so H_N^* increases with the population size.

Chapter 5

On selection in finite populations

Chai Molina and David J. D. Earn

5.1 Abstract

Two of the major forces shaping evolution are drift and selection. Although there are a variety of models of neutral drift, only a few—mostly based on the Wright-Fisher and Moran processes—extend to situations in which selection and drift act together on a finite population. These models are not applicable to all biological populations, and even models of neutral drift can display behaviour very different from that of the neutral Wright-Fisher and Moran processes. Previous studies of a different class of models of selection made assumptions that are useful in obtaining continuum limits, fixation times and other quantities of interest to population geneticists, but exclude the Wright-Fisher model except in the continuum limit. In addition, analyses of evolutionary stability in finite populations depend only on fixation probabilities, which can be evaluated under less restrictive assumptions than those required to estimate fixation times or other more complex population-genetic quantities. We therefore make fewer assumptions and define a selection process more broadly to be any member of a large class of finite-population, mutationless models of selection and drift (which include the Wright-Fisher and Moran processes as special cases). We derive an intuitive criterion for selection favouring fixation of one strategy over another for *any* selection process. Applied to evolutionary games played in finite populations, this criterion yields sufficient conditions for the evolutionary robustness and stability of a strategy under *any* selection process.

5.2 Introduction

Two key determinants of the distribution of traits in a population are ***drift*** (stochasticity in the temporal evolution of trait frequencies in finite populations) and ***selection*** (the process by which traits associated with higher fitness—*i.e.,* greater expected lifetime reproductive output—increase in frequency over time [60, 61, 62]). There are many mathematical models of neutral drift—when no variability in fitness is associated with the evolving traits [63, 64, 65, 69, 202, 203, 204, 205]—but few that extend to traits involving variable fitness. In fact, almost all models in the literature involving both selection and drift are variations of the classical Moran [63] and Wright-Fisher [64, 65] (WF) processes (described in §§5.4.1 and 5.4.2 below).

Even in the case of neutral drift, other models can behave very differently from the Moran and WF processes [67, 69, 70, 71, 72, 73, 74]. Motivated by this, and by the fact that not all biological populations satisfy the assumptions of the Moran and WF models relating to the mode of reproduction (*e.g.,* Pacific Oysters [66, 67]), Der and co-workers [68, 75] defined and analyzed Generalized Wright-Fisher (GWF) models (which include the Eldon-Wakeley process [67, 206]). They showed that fixation probabilities, as well as other population-genetic quantities of interest, can vary substantially if the assumptions of the WF model are relaxed. Moreover, fitting alternative models of selection to empirical data on the dynamics of allele frequencies in fruit flies suggests that the alternative models have at least as much explanatory power as the WF model [68]. Greater understanding of more general selection processes in finite populations would be valuable.

The Moran and WF models have also recently been used to develop evolutionary game theory. In finite populations, strategies that yield lower mean payoffs (*e.g.,* deleterious mutations) can have positive fixation probabilities, so evolutionarily stable strategies (ESSs) should be defined to be resistant to both invasion and fixation [186] (see definition 5.6.3). Which strategies turn out to be ESSs may depend on the selection process: it has been shown that different "updating rules" (*i.e.,* the various processes by which variability in fitness can influence the frequencies of strategies in the population) can yield different evolutionary dynamics [76]. However, almost all results pertaining to evolutionary stability in finite populations (and fixation probabilities in particular) obtained thus far have been based on either the Moran [186, 193, 207, 208, 209] or WF [81, 193] processes. One exception is the analysis of a Cannings exchangeable allele model (see [202]) modified to include selection, which is, however, limited by the assumption of weak selection (as are many other studies applying only to the Moran or WF models). A promising approach to accommodating selection processes other than the WF and Moran models in evolutionary game theory consists of a framework for analyzing games with discrete strategies, a positive mutation rate (identical for all strategies), and an arbitrary updating rule, in the limit of weak selection [77, 78]. This approach has been extended to continuous strategy sets with small mutations and continuous time [79], in which case the assumption of weak selection can be relaxed. While these studies supply a useful framework in which to work, they

involve calculating parameters that depend on the updating scheme and population structure (but independent of the game) in order to characterize when one strategy is favoured over another. This drawback may make results that are robust to the choice of selection process harder to obtain.

A general theory of the population-level processes of drift and selection will promote progress in both population genetics and evolutionary game theory. Applications in evolutionary game theory often involve fixation probabilities only. It is therefore useful to relax some of the assumptions of the framework of GWF models, which facilitate analysis of continuum-limits and more complex population-genetic quantities such as fixation times [68, 75].

Here, we define a large class of biologically sensible models of selection in finite populations (which contains the class of GWF models), and a subclass of models of neutral drift. We study the probability of fixation of traits under these models and obtain an intuitive result whereby traits yielding a higher fitness regardless of their frequency in the population are more likely to fix than traits that do not confer a selective advantage. We then apply this result in the context of evolutionary games in finite populations, in which both the game payoffs and the fitnesses of individuals with a given payoff are stochastic. To our knowledge, these are the first results in evolutionary game theory that apply to n-player discrete-strategy games (for any $n \geq 2$) and are robust to any of the particular details of drift and selection (as well as entirely independent of the intensity of selection).

5.3 General selection processes

Consider an asexual population of N agents comprised of two phenotypes, A and B. Let $\overline{W}_A(i)$ and $\overline{W}_B(i)$ be the mean fitnesses of agents of type A and B, respectively, when there are i agents ($1 \leq i \leq N-1$) of type A in the population[1]. For discrete times $t \in \mathbb{N} = \{0, 1, 2, \ldots\}$, let $X(t)$ be the number of agents of type A at time t. We refer to $X(t)$ as the ***state*** of the population at time t, and to $X(0)$ as the ***initial state*** of the population.

Suppose that the population size remains constant and equal to N and that the population composition evolves according to a discrete-time Markov process with a stationary transition matrix P: the probability of the population state at time $t+1$ being $X(t+1) = j$ is dependent only on the population state $X(t)$ at time t (but not on the time t itself), and

$$P_{i,j} = \Pr\left(X(t+1) = j | X(t) = i\right). \tag{5.1}$$

$P = (P_{i,j})$ is a stochastic matrix, that is, $P_{i,j} \geq 0$ and $\sum_{j=0}^{N} P_{i,j} = 1$ for all i, $0 \leq i \leq$

[1]Fitnesses need not be defined for $i = 0$ or N, as in these extremes the population is homogeneous and there is no variability in fitness.

N. For example, the frequency dependent Moran and Wright-Fisher processes [61, 62] specify how to construct the transition matrix $P_{i,j}$ from the fitnesses $\overline{W}_A(i)$ and $\overline{W}_B(i)$ (see §§5.4.1 and 5.4.2 below).

We assume that there are no mutations, which also implies that if the entire population is composed of one type (A or B), then it will remain at that state forever (that is to say, the states in which the population is monomorphic are ***absorbing***). By a ***mixed-type state*** we mean a population of As and Bs including at least one of each type.

Definition 5.3.1. *We say that P defines a (mutationless) **selection process** with respect to the mean fitnesses $\overline{W}_A(i)$ and $\overline{W}_B(i)$ ($1 \leq i \leq N-1$) if it satisfies the following biologically sensible properties:*

*H1 At any state $X(t) = i$, the fitness of individuals of one type is higher than that of the other, if and only if (*iff*) the expected number of individuals of the type having higher fitness in the next time step ($t+1$) is higher than their number at time t. Mathematically, for $1 \leq i \leq N-1$,*

$$\overline{W}_A(i) > \overline{W}_B(i) \quad \Longleftrightarrow \quad \mathbb{E}\left(X(t+1)|X(t)=i\right) = \sum_{j=0}^{N} jP_{i,j} > i = X(t)\,, \tag{5.2a}$$

and

$$\overline{W}_B(i) > \overline{W}_A(i) \quad \Longleftrightarrow \quad \mathbb{E}\left(X(t+1)|X(t)=i\right) = \sum_{j=0}^{N} jP_{i,j} < i = X(t)\,. \tag{5.2b}$$

*H2 If at time τ, both types are present in the population (that is, the population is at a mixed-type state), then there is a positive probability of the population becoming monomorphic (*i.e.*, reaching state 0 or N) in finite time. That is, for all $1 \leq i \leq N-1$, there exists $t > \tau$ (possibly dependent on i) such that*

$$\Pr\left(X(t) = 0 \text{ or } X(t) = N | X(\tau) = i\right) > 0\,. \tag{5.3}$$

H3 The states 0 and N are absorbing, that is, once reached, the population remains there forever: for all $\tau \geq 0$ and $t \geq \tau$,

$$\Pr\left(X(t) = 0 | X(\tau) = 0\right) = 1\,, \tag{5.4a}$$
$$\Pr\left(X(t) = N | X(\tau) = N\right) = 1\,. \tag{5.4b}$$

Remark 5.3.2. *In this article, we analyze only selection processes without mutation; see [77] for an analysis of selection processes that include mutation (at equal rates for all*

types, in the limit of weak selection).

We will find the following definition from the theory of Markov processes useful:

Definition 5.3.3. *We say that state j is* ***accessible*** *from state i (or that state i* ***leads*** *to state j) if, starting from state $X(0) = i$ it is possible to arrive at state j in finite time,* i.e., *there is a time $\tau \geq 0$ such that* $\Pr\left(X(\tau) = j | X(0) = i\right) > 0$.

Remark 5.3.4. *Equivalently, the state j is accessible from state i* iff *there exists $n \geq 1$ such that $(P^n)_{i,j} > 0$.*

Some selection processes (*e.g.,* the Moran and WF processes; see §§5.4.1 and 5.4.2) have an additional property, which is not strictly necessary for the analysis that follows, but is biologically sensible and simplifies some of the statements of our results:

Definition 5.3.5. *We say that a selection process is* ***mixed-irreducible*** *if any two mixed-type states are accessible from one another.*

A process being mixed-irreducible does not imply that the transition matrix P is an irreducible matrix. In fact, P cannot be irreducible because of the absorbing homogeneous states. However, the submatrix corresponding to the non-homogeneous (mixed-type) states ($\widetilde{P} = (P_{i,j})_{i,j=1}^{N-1}$) must be irreducible. Equivalently, a selection process is mixed-irreducible if and only if for any mixed-type states, $1 \leq i \leq N-1$ and $1 \leq j \leq N-1$, there is a time $\tau_{i,j} > 0$ such that

$$\Pr\left(X(t + \tau_{i,j}) = j | X(t) = i\right) > 0\,. \tag{5.5}$$

Using definition 5.3.3, H2 can be restated as follows: any state $1 \leq i \leq N-1$ leads to 0 or N. However, by a standard result in the theory of Markov processes, it is not only possible, but certain, that the process reaches one of the absorbing states in finite time:

Proposition 5.3.6. *A selection process reaches one of the absorbing states, 0 or N, in finite time: for any $0 \leq i \leq N$,*

$$\Pr\left(\exists t \in \mathbb{N} \textit{ such that } X(t) \in \{0, N\} | X(0) = i\right) = 1. \tag{5.6}$$

Proof. If $X(0) = 0$ or $X(0) = 0$, nothing remains to be shown.

Let $C = \{1, 2, \ldots, N-1\}$ and consider $i \in C$. Suppose, in order to derive a contradiction, that the absorption probability starting from state i is

$$\Pr\left(\exists t \in \mathbb{N} \text{ such that } X(t) \in \{0, N\} | X(0) = i\right) < 1. \tag{5.7}$$

Then,

$$\Pr\left(X(t) \in C \text{ for all } t \in \mathbb{N} | X(0) = i\right) > 0. \tag{5.8}$$

If $X(t)$ takes values in C for all times $t \geq 0$, then since C is finite, at least one index

$0 \leq j \leq N-1$ is visited infinitely often, that is, for some $1 \leq j \leq N-1$,

$$\Pr\left(\text{for any } T \geq 0, \text{ there exists } t > T \text{ such that } X(t) = j | X(0) = i\right) > 0\,. \tag{5.9}$$

Now note that H1 implies that C is a set of inessential, and therefore nonrecurrent states (see appendix 5.A.2 and [210, theorem I.4.4]), which cannot be visited infinitely often [210, theorem I.4.3] contradicting equation (5.9). □

Hence we can define the following:

Definition 5.3.7 (Fixation time and probabilities)**.** *For any mutationless selection process,*

1. *the first time at which the process arrives at one of the absorbing states is the* ***fixation time****, that is,* $T_{\text{fix}} = \min\{t | X(t) = 0 \text{ or } N\}$.
2. *for any* $0 \leq i \leq N$, $p_{\text{fix}}(i)$ *is the probability of reaching the absorbing state* N, i.e., ***fixation probability of*** A *from the initial state* i*:*

$$p_{\text{fix}}(i) = \Pr\left(\lim_{t\to\infty} X(t) = N | X(0) = i\right)\,. \tag{5.10}$$

Because fixation is assured (proposition 5.3.6), the probability of fixation of B starting from state i (defined similarly) is $1 - p_{\text{fix}}(i)$. Note that since the states $X = N$ and $X = 0$ are absorbing, $p_{\text{fix}}(0) = 0$ and $p_{\text{fix}}(N) = 1$. Also, proposition 5.3.6 implies that the fixation time T_{fix} is a non-negative random variable satisfying $\Pr(T_{\text{fix}} < \infty) = 1$.

Intuitively, under neutral drift (absense of selection), the expected number of individuals of each type at time $t+1$ should be equal to their numbers at time t, that is, if $X(t) = i$, then $\mathbb{E}\left(X(t+1)\right) = i = X(t)$. This motivates the following:

Definition 5.3.8. *We say that the transition matrix* P *defines a* ***neutral drift process*** *if* $X(t)$ *satisfies H2, H3 and*

$$\mathbb{E}\left(X(t+1)\right) = X(t)\,. \tag{5.11}$$

Alternatively, we say that $X(t)$ *is a neutral drift process.*

Remark 5.3.9. P *defines a neutral drift process* iff *for any* $0 \leq i \leq N$, $\sum_{j=0}^{N} j P_{i,j} = i$.

Since $X(t)$ is a bounded Markov process, if P defines a neutral drift process, equation (5.11) implies that $X(t)$ is also a martingale (see definition 5.A.2).

5.4 Particular selection processes

In this section, we discuss three population processes from the literature and establish that they are selection or neutral drift processes according to definitions 5.3.1 and 5.3.8. This

amounts to verifying H2, H3 and either H1 or equation (5.11).

§§5.4.1 and 5.4.2 show that the frequency-dependent Moran and Wright-Fisher processes are mixed-irreducible selection processes (definition 5.3.5). Moreover, when the fitnesses of types A and B are equal, both are neutral drift processes. § 5.4.3 presents another selection process from the literature.

5.4.1 The Moran process

If the population evolves according to the Moran process [61, 62, 63], then exactly one individual is replaced at each time step. In detail, at each time step:

- an agent is chosen for death, with equal probability for all agents;
- an agent is chosen for reproduction, with probability proportional to its fitness[2];
- the agent chosen for death is replaced with a clone of the agent chosen for reproduction.

Note that sampling of agents is done with replacement, so that an agent can be chosen for both death and reproduction (in which case the population remains unchanged).

When the population consists of i mutants (individuals of type A) and $N-i$ residents (individuals of type B), the probabilities of choosing a mutant or a resident for death are i/N and $(N-i)/N$, respectively. The probabilities of choosing a mutant or a resident for reproduction are

$$\frac{i\overline{W}_A(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\,, \tag{5.12a}$$

and

$$\frac{(N-i)\overline{W}_B(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\,. \tag{5.12b}$$

Because the death and reproduction events are independent, the transition probabilities are simply

$$P_{i,i+1}=\frac{i\overline{W}_A(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\times\frac{N-i}{N}>0\,, \tag{5.13a}$$

$$P_{i,i-1}=\frac{(N-i)\overline{W}_B(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\times\frac{i}{N}>0\,, \tag{5.13b}$$

[2]We assume here that the fitnesses $\overline{W}_A(j)$ and $\overline{W}_B(j)$ are positive for $1\leq j\leq N-1$.

and (since at each time step at most one individual is replaced)

$$P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1} = \frac{i^2\overline{W}_A(i) + (N-i)^2\overline{W}_B(i)}{N\big(i\overline{W}_A(i) + (N-i)\overline{W}_B(i)\big)} > 0\,. \tag{5.13c}$$

Lastly, $P_{0,0} = P_{N,N} = 1$ and $P_{0,i} = P_{N,N-i} = 0$ for all $1 \leq i \leq N$ (the states where the resident or mutant have fixed are absorbing, so H3 is trivially satisfied).

For any $1 \leq i \leq N-1$, if $X(t) = i$, we have

$$\begin{aligned}\mathbb{E}\big(X(t+1)\big) - X(t) &= -i + \sum_{j=0}^{N} jP_{i,j} = -i + \big[(i-1)P_{i,i-1} + iP_{i,i} + (i+1)P_{i,i+1}\big] \\ &= -i + \big[i + P_{i,i+1} - P_{i,i-1}\big] = \frac{i(N-i)\big(\overline{W}_A(i) - \overline{W}_B(i)\big)}{N\big(i\overline{W}_A(i) + (N-i)\overline{W}_B(i)\big)}\,.\end{aligned} \tag{5.14}$$

The expected number of individuals of type A (respectively B) in the next time-step is larger than in the current time-step, if and only if $\overline{W}_A(i) > \overline{W}_B(i)$ (respectively $\overline{W}_B(i) > \overline{W}_A(i)$), so H1 is satisfied.

To see that H2 is satisfied, and moreover, that P defines a mixed-irreducible selection process, consider i and j such that $1 \leq i \leq N-1$, $0 \leq j \leq N$ and $j \neq i$, and observe that there is a positive probability of changing from state i to state j in $d = |j-i|$ steps: setting $\sigma = \operatorname{sign}(j-i)$, we have

$$\Pr\big(X(t+d) = j | X(t) = i\big) = \prod_{k=1}^{d} P_{(i+\sigma(k-1)),(i+\sigma k)} > 0 \tag{5.15a}$$

$$\Pr\big(X(t+1) = i | X(t) = i\big) = P_{i,i} > 0\,, \tag{5.15b}$$

so all states can be reached from state $X(t) = i$ in finite time, and in particular, the probability of B fixing at a future time $t+\tau$ ($\tau \geq 0$) is positive.

If neither type has a selective advantage over the other, regardless of their frequencies in the population, then for all $1 \leq i \leq N-1$, $\overline{W}_A(i) = \overline{W}_B(i)$, so from equation (5.14), $\mathbb{E}\big(X(t+1)\big) = X(t)$, and P defines a neutral drift process.

5.4.2 The Wright-Fisher process

If the population evolves according to the Wright-Fisher process [61, 62] then all individuals are replaced at each time step (generations do not overlap). At each time step, the entire population of N individuals is replaced by a new generation constructed using bi-

nomial sampling: in each of the N Bernoulli trials, the probability of drawing any type represented in the current generation is proportional to its present mean fitness and to the present number of individuals of that type. Thus, the probability that an individual in the next generation will be of type A is

$$\frac{i\overline{W}_A(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}, \tag{5.16}$$

and

$$\begin{aligned} P_{i,j} &= \Pr\left(X(t+1)=j|X(t)=i\right) \\ &= \binom{N}{j}\left(\frac{i\overline{W}_A(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\right)^j\left(\frac{(N-i)\overline{W}_B(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}\right)^{N-j}, \end{aligned} \tag{5.17}$$

where $P_{0,0}=P_{N,N}=1$ (so the states $X=0$ and $X=N$ are absorbing and H3 is satisfied). Note that if A is not present at some time τ, B has fixed and the population remains at state $X(t)=0$ for all $t\geq\tau$, and similarly if B is not present at some time τ, then $X(t)=N$ for all $t\geq\tau$.

The mean of a binomial random variable defined by n trials with success probability p is np, so for any $0\leq i\leq N$, we have

$$\begin{aligned} \mathbb{E}\left(X(t+1|X(t)=i)\right)-i &= N\frac{i\overline{W}_A(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}-i \\ &= i(N-i)\frac{\overline{W}_A(i)-\overline{W}_B(i)}{i\overline{W}_A(i)+(N-i)\overline{W}_B(i)}, \end{aligned} \tag{5.18}$$

so H1 is satisfied. H2 is trivially satisfied because for any $1\leq i\leq N-1$, $P_{i,0}>0$. Thus, P defines a selection process, which is, moreover, mixed-irreducible, because for any $1\leq i\leq N-1$, $P_{i,j}>0$ also for any $1\leq j\leq N-1$.

If neither type has a selective advantage over the other, $\overline{W}_A(i)=\overline{W}_B(i)$ for all $1\leq i\leq N-1$, and equation (5.18) becomes $\mathbb{E}\left(X(t+1|X(t)=i)\right)=i=X(t)$, so $X(t)$ is a neutral drift process.

5.4.3 The Eldon-Wakeley process with viability selection

The Eldon-Wakeley (EW) process [67, 206] is a variation on the neutral Moran process that allows for a skewed (rather than uniform) offspring distribution. It has been used to interpret genetic data from Pacific Oysters [67, 206].

The EW process describes neutral drift in a population of constant size N, consisting

of two types, A and B. At each time step, a single agent is randomly drawn from the population with uniform probability, and produces a random number of offspring $U-1$. The parent agent survives to the next generation and its $U-1$ offspring replace $U-1$ randomly chosen members of the remainder of the population. In the special case that exactly one offspring is always produced, *i.e.,* $\Pr(U=2)=1$, the EW process is similar (but not identical) to the classical Moran process [62, 63]: in both processes, the parent always produces one offspring, which increases the number of individuals of the parent's type in the next generation *iff* the agent chosen to be replaced is not of the parent's type. In the EW process, the parent is guaranteed to survive, and one additional offspring replaces another randomly chosen member of the population, so if there are i agents of type A, the probability that the population state remains the same is

$$\frac{i}{N}\frac{i-1}{N-1}+\frac{N-i}{N}\frac{N-i-1}{N-1}=\frac{i^2+(N-i)^2-N}{N(N-1)}\,. \tag{5.19}$$

By contrast, in the Moran process, this probability is given by $\frac{i^2+(N-i)^2}{N^2}$(see equation (5.13c)). Thus, whenever the population is at a mixed-type state (*i.e.,* $1\leq i\leq N-1$), the probability that the population state remains unchanged is larger for the Moran model than for the EW model. However, for both models, the probability of increase in type A is the same as the probability of increase in type B (this probability does depend on the population composition). Thus, in effect, the neutral (*i.e.,* selectionless) EW process with $U=2$ is a slightly "sped up" version of the neutral Moran process, where only time steps in which the population state is changed are counted.[3]

Letting $X(t)=i$ be the number of individuals of type A at some time $t\geq 0$, then the probabilities that an agent of type A and B are chosen for reproduction are i/N and $(N-i)/N$, respectively. If an agent of type A is chosen for reproduction and produces $U-1=u-1$ offspring, then the distribution of the number of B agents chosen for replacement is hypergeometrically distributed with sample size $N-1$, initial configuration $N-i$ and $u-1$ draws [206], so the probability of k agents of type B ($0\leq k\leq u-1$) being replaced by As is

$$\frac{\binom{N-i}{k}\binom{i-1}{u-1-k}}{\binom{N-1}{u-1}}\,, \tag{5.20}$$

which has mean $(u-1)\frac{N-i}{N-1}$. Similarly, the mean number of agents of type A to be replaced, given that a B agent is chosen for reproduction and produces $u-1$ offspring is $(u-1)\frac{i}{N-1}$. Thus, by the law of total expectation (theorem 5.A.1, conditioning on the type of

[3] In the original version of the EW process [67], the fitnesses of both types were equal, and the parent agent was guaranteed to survive to the next generation. Der and coworkers [206] generalized the original model to types with different fitnesses, but in their version of the EW process, it is possible for the parent to be chosen for replacement. Here, we reformulate Der *et al.*'s extended model while retaining the original condition that the reproducing agent cannot be chosen for replacement. In contrast to our version of the EW process, setting $U=2$ in Der *et al.*'s version yields the Moran process exactly.

agent chosen for reproduction), the expected number of individuals of type A in the next generation, given their present number, is:

$$\begin{aligned}\mathbb{E}\left(X(t+1)|X(t)=i\right) &= \frac{i}{N}\,\mathbb{E}\left(i+(U-1)\frac{N-i}{N-1}\right)+\frac{N-i}{N}\,\mathbb{E}\left(i-(U-1)\frac{i}{N-1}\right)\\ &= i+\frac{i}{N}\frac{N-i}{N-1}\,\mathbb{E}\,(U-1)-\frac{N-i}{N}\frac{i}{N-1}\,\mathbb{E}\,(U-1)\\ &= i\,. \end{aligned} \tag{5.21}$$

Der *et al.* [206] have generalized the neutral EW process [67] by adding a deterministic "viability selection" step: for $s \in \mathbb{R}$, given the population state $X(t)$ at time t, an intermediate, pre-selection offspring population state at time $t+1$ is generated according to the EW model without selection (described above). The population state $X(t+1)$ at time $t+1$ is then obtained by transforming the pre-selection offspring state according to standard (deterministic) logistic growth:

$$i \mapsto v\,(i) = \left\lfloor \frac{(1+s/N)i}{(1+s/N)i+(N-i)}N \right\rfloor = \left\lfloor \frac{N+s}{N+s(i/N)}i \right\rfloor, \tag{5.22}$$

where $\lfloor x \rfloor$ is the largest integer smaller than x. This corresponds to selection acting on the offspring before reaching reproductive age ($X(t)$ represents the state of the reproductively-mature population).

Now observe that for any $1 \le i \le N-1$, if $s > 0$ then

$$v\,(i) \ge i\,, \tag{5.23}$$

if $s < 0$

$$v\,(i) \le i\,, \tag{5.24}$$

and if $s = 0$, $v\,(i) = i$ (so the original EW process is recovered). Note also that because $\frac{(1+s/N)i}{(1+s/N)i+(N-i)}N < N$, fixation cannot occur in the selection step.

For any s, the selection step and neutral EW process above define a Markov process. Equations (5.21) and (5.23) imply that H1 is satisfied for this Markov process.

To verify H2 for any $s \ge 0$, choose any i ($1 \le i \le N-1$) and $u \ge 2$ such that $\Pr(U=u) = p_u > 0$ (such u must exist because otherwise no offspring are ever created). The probability of an individual of type A reproducing is $\frac{i}{N}$. Using equations (5.20), the probability of increasing the number of As in the population given that an individual of type A reproduces and that $U=u$ is

$$p_+(i) = 1 - \frac{\binom{N-i}{0}\binom{i-1}{u}}{\binom{N-1}{u}}\,, \tag{5.25}$$

and $p_+(i) > 0$ because $i < N$. Hence, the probability of increasing the number of agents of type A in the population in the next generation is no less than

$$\Pr\big(X(t+1) > i | X(t) = i\big) \geq \frac{i}{N} p_u p_+(i) > 0\,, \tag{5.26}$$

(recall that the selection step cannot decrease the number of As in the population; see equation (5.23)). Now, starting from state i, if the number of agents of type A in the population is increased at each step, fixation of A is attained in at most $N - i$ steps. Since the probability of increasing the number of A's in the population is positive for $1 \leq i < N$, the probability of A fixing in i steps is positive,

$$\Pr\big(X(t+i) = N | X(t) = i\big) > 0\,. \tag{5.27}$$

Verifying H2 for $s < 0$ is similar.

As in §§5.4.1 and 5.4.2, H3 is satisfied because there is no mutation, and consequently the EW process with viability selection defines a selection process.

Note that equation (5.21) implies that in the absence of selection, the EW process is a neutral drift process. Moreover, a similar method to that used in §5.4.1 shows that the EW process without selection is mixed-irreducible.

5.5 Bounds on fixation probabilities

We begin by calculating the fixation probabilities $p_{\text{fix}}(i)$ for a neutral drift process, generalizing [68, Theorem 2]:

Proposition 5.5.1 (fixation under neutral drift)**.** *If $X(t)$ is a neutral drift process, then for any $0 \leq i \leq N$, if $X(0) = i$, the fixation probability of A is $p_{\text{fix}}(i) = i/N$.*

Proof. Define the random variable $T_A = \min\{t | X(t) = N\}$, that is, the fixation time of A ($T_A = \infty$ if A never fixes). Similarly, let $T_B = \min\{t | X(t) = 0\}$ be the fixation time of B. Both T_A and T_B are stopping times (see definition 5.A.3), and hence the fixation time $T_{\text{fix}} = \min\{T_A, T_B\}$ is also a stopping time [80, p.256]. Since either A or B must fix, $\Pr(T_{\text{fix}} < \infty) = 1$ (proposition 5.3.6). For any t, we have $0 \leq X(t) \leq N$, so it follows that for all $t \geq 0$,

$$\mathbb{E}\left(\sup_{t \geq 0} X(\min\{T, t\})\right) < \infty\,. \tag{5.28}$$

Thus, since $X(t)$ is a martingale, the optional stopping theorem (theorem 5.A.5) implies that

$$X(0) = \mathbb{E}\big(X(0)\big) = \mathbb{E}\big(X(T_{\text{fix}})\big) = \Pr\big(X(T_{\text{fix}}) = 0\big) \cdot 0 + \Pr\big(X(T_{\text{fix}}) = N\big) \cdot N\,, \tag{5.29}$$

and

$$p_{\text{fix}}(i) = \Pr\left(\lim_{n\to\infty} X(t) = N | X(0) = i\right) = \Pr\left(X(T_{\text{fix}}) = N | X(0) = i\right) = i/N\,. \tag{5.30}$$

□

Remark 5.5.2. *For an intuitive explanation of proposition 5.5.1, consider that if the population consists of N equally fit types (instead of two) a symmetry argument shows that all types are equally likely to fix. If fixation is assured, then each type fixes with probability $1/N$.*

Now return to the scenario of only two segregating types. If initially (at time $t = 0$) there are no individuals of type A then A cannot fix (because we assume no mutation), so $p_{\text{fix}}(0) = 0$; similarly, $p_{\text{fix}}(N) = 1$.

If the initial number of individuals of type A is $1 \leq X(0) = i \leq N - 1$, label these as individuals $1, \ldots, i$, and label the individuals of type B as $i+1, \ldots, N$, so that all individuals are distinguishable. Define a heritable "supertype" *as both the individual label, and the previously defined trait, A or B (*e.g., *individual 1 is now of type $(1, A)$, and individual $i+1$ is now of type $(i+1, B)$). With this new definition, there are now N different supertypes segregating in the population: for $1 \leq j \leq i$, the descendants of an individual of supertype (j, A) are also of type (j, A), and for $i + 1 \leq j \leq N$ the descendants of an individual of type (j, B) are of type (j, B). If neither type A or B has a selective advantage, then the fixation probability of each supertype is $1/N$. The fixation probability of type A is then the sum of the fixation probabilities of supertypes (j, A) for $1 \leq j \leq i$, that is i/N.*

Proposition 5.5.1 shows that fixation probabilities are identical for all neutral drift processes. Thus, fixation probabilities under neutral drift can be used as a baseline for comparing fixation probabilities under selection, motivating the following definition of selection favouring or opposing fixation of an invading mutant :

Definition 5.5.3. *If there are i agents of type A and $N - i$ agents of type B in a population undergoing selection, we say that* ***selection favours fixation*** *of A if the probability of A fixing is $p_{\text{fix}}(i) > i/N$, and* ***selection opposes fixation*** *of A if $p_{\text{fix}}(i) < i/N$.*

Lemma 5.5.4 below gives intuitive sufficient conditions for selection opposing fixation: if type A is never fitter than type B, and is less fit at some state that is accessible from the initial one, then selection opposes fixation of A.

Lemma 5.5.4 (Sufficient conditions for selection opposing fixation)**.** *Consider a population of constant size N in which there are two types, A and B, evolving under a selection process $\mathscr{P}$. Let $\overline{W}_A(i)$ and $\overline{W}_B(i)$ be the mean fitnesses of types A and B (respectively) when there are i individuals of type A in the population, and let $\mathcal{S}_i$ be the set of mixed-type states that are accessible from state i under $\mathscr{P}$ (so $\mathcal{S}_i \subset \{1, 2, \ldots, N-1\}$).*

If $X(0) = i$ denotes the initial state $(0 \leq i \leq N)$, and an individual of type A is no

fitter than *an individual of type B at any population state $j \in \mathcal{S}_i$,* i.e., *if*

$$\overline{W}_A(j) \leq \overline{W}_B(j) \,, \qquad \textit{for each } j \in \mathcal{S}_i \,, \tag{5.31}$$

then the probabilities of A and B fixing satisfy $p_{\text{fix}}(i) \leq i/N$ and $1 - p_{\text{fix}}(i) \geq (N-i)/N$, respectively.

If, in addition, there exists a state $\hat{\imath} \in \mathcal{S}_i$ at which type A is strictly less fit than type B, i.e.,

$$\overline{W}_A(\hat{\imath}) < \overline{W}_B(\hat{\imath}) \,, \qquad \textit{for some } \hat{\imath} \in \mathcal{S}_i \,, \tag{5.32}$$

then selection opposes fixation of A, i.e., *the probability of A fixing is* strictly less *than under neutral drift ($p_{\text{fix}}(i) < i/N$) and the probability of B fixing is* strictly greater *than under neutral drift ($1 - p_{\text{fix}}(i) > (N-i)/N$).*

Proof. Observe that $X(t)$ is a non-negative supermartingale (definition 5.A.2). Thus, for any stopping time S, with $\Pr(S < \infty) = 1$, a version of the optional stopping theorem for supermartingales (theorem 5.A.6) states that

$$\mathbb{E}\left(X(S)\right) \leq \mathbb{E}\left(X(0)\right). \tag{5.33}$$

Using a constant stopping time $S = \tau \geq 0$, equation (5.33) gives

$$\mathbb{E}\left(X(\tau)|X(0) = i\right) \leq X(0) = i \,. \tag{5.34}$$

Letting T_{fix} be the fixation time for the system, by proposition 5.3.6 we can apply equation (5.33) to show that for any initial state $X(0) = i$ for ($0 \leq i \leq N$) the fixation probability of A satisfies

$$p_{\text{fix}}(i)\, N = \mathbb{E}\left(X(T_{\text{fix}})|X(0) = i\right) \leq X(0) = i \,, \tag{5.35}$$

so $p_{\text{fix}}(i) \leq i/N$, and the fixation probability of B is $1 - p_{\text{fix}}(i) \leq (N-i)/N$.

Similarly (using H1 as well)

$$p_{\text{fix}}(\hat{\imath})\, N = \mathbb{E}\left(X(T_{\text{fix}})|X(0) = \hat{\imath}\right) \leq \mathbb{E}\left(X(1)|X(0) = \hat{\imath}\right) < \hat{\imath} \,, \tag{5.36}$$

so $p_{\text{fix}}(\hat{\imath}) < \hat{\imath}/N$.

Denoting the probability of reaching state j at time $\tau \geq 0$ starting from state $X(0) = i$ by

$$P_{i,j}^{(\tau)} = \Pr\left(X(\tau) = j|X(0) = i\right), \tag{5.37}$$

we have $P_{i,j}^{(\tau)} = (P^\tau)_{i,j}$.

If i leads to $\hat{\imath}$, then for some time $\tau \geq 0$, the probability of reaching state $\hat{\imath}$ from state i

is nonzero, $P_{i,\hat{\imath}}^{(\tau)} > 0$. Conditioning on the state arrived at in the τ-th time-step, we have

$$p_{\text{fix}}(i) = \sum_{j=0}^{j=N} P_{i,j}^{(\tau)} p_{\text{fix}}(j) = \sum_{\substack{j=0 \\ j\neq\hat{\imath}}}^{j=N} P_{i,j}^{(\tau)} p_{\text{fix}}(j) + P_{i,\hat{\imath}}^{(\tau)} p_{\text{fix}}(\hat{\imath}) < \frac{1}{N} \sum_{j=0}^{j=N} P_{i,j}^{(\tau)} j$$
$$= \frac{1}{N} \mathbb{E}\left(X(\tau) | X(0) = i\right). \tag{5.38}$$

Using equation (5.34), we obtain

$$p_{\text{fix}}(i) < \frac{1}{N} X(0) = \frac{i}{N}. \tag{5.39}$$

and the probability of B fixing is $1 - p_{\text{fix}}(i) > \frac{N-i}{N}$. □

Note that Refs. [211, 212] found (without defining a selection process) that the fixation probability of a selectively advantageous mutation is *no less than* that of a neutral one. For a general selection process, we have identified and rigorously established conditions under which a selectively advantageous mutation fixes with probability *strictly larger* than neutral.

Under the hypotheses of lemma 5.5.4, if the state $\hat{\imath}$ at which the A agents' fitness is lower than that of B agents is accessible from any other state (for the selection process in question), then $p_{\text{fix}}(i) < i/N$ for all $1 \leq i \leq N-1$. It follows that:

Corollary 5.5.5. *If the hypotheses of lemma 5.5.4 hold, and the selection process is mixed-irreducible, then for any mixed-type initial state $1 \leq i \leq N-1$, $p_{\text{fix}}(i) < i/N$ so selection opposes fixation of A.*

Corollary 5.5.5 generalizes [81, Theorem 1], which applies only to the Wright-Fisher process. While the proof given in [81] is easily extended to arbitrary mixed-irreducible selection processes, the proof of lemma 5.5.4 given above is both more general, and renders the biological mechanism responsible for the reduced fixation probability compared to neutral drift processes more transparent: Under neutral drift processes, the mean number of individuals of each type *does not change* from one time step to the next. By contrast, under the conditions of lemma 5.5.4, H1 implies only that the mean number of agents of type A *does not increase* over time. Moreover, if the process is at the state $\hat{\imath}$ (at which A is less fit), then the mean number of agents of type A decreases in the next generation. Because $\hat{\imath}$ is accessible from the initial population state, this increases the probability that A decreases in frequency over time (compared to neutral drift processes), which translates to a lower fixation probability.

Lemma 5.5.6 below is a partial converse to lemma 5.5.4; together, lemmas 5.5.4 and 5.5.6 show that equations (5.41) and (5.42) characterize the situations in which selection opposes fixation irrespective of the selection process.

Lemma 5.5.6 (Necessary conditions for selection opposing fixation for any selection process)**.** *Consider a population of constant size N in which there are two types, A and B. Let $\overline{W}_A(i)$ and $\overline{W}_B(i)$ be the mean fitnesses of types A and B (respectively) when there are i individuals of type A in the population.*

Suppose that the population's initial state is $X(0) = i$ $(0 \leq i \leq N)$ and, for any selection process, selection opposes fixation of A, that is,

$$p_{\text{fix}}(i) < i/N\,, \qquad \textit{for any selection process}\,. \tag{5.40}$$

Then:

- *the mean fitness of an individual of type A is* no larger *than that of an individual of type B at any mixed-type state,* i.e.,

$$\overline{W}_A(j) \leq \overline{W}_B(j)\,, \qquad \textit{for all } j, \quad 1 \leq j \leq N-1\,, \tag{5.41}$$

- *there exists a mixed-type state at which the mean fitness of type A is smaller than type B,* i.e.,

$$\overline{W}_A(\hat{\imath}) < \overline{W}_B(\hat{\imath})\,, \qquad \textit{for some } \hat{\imath}, \quad 1 \leq \hat{\imath} \leq N-1\,. \tag{5.42}$$

Proof. Suppose, in order to derive a contradiction, that equation (5.42) does not hold: for all states $1 \leq j \leq N-1$

$$\overline{W}_A(j) \geq \overline{W}_B(j)\,. \tag{5.43}$$

Then for any selection process, from lemma 5.5.4 (with the roles of A and B reversed), $p_{\text{fix}}(i) \geq i/N$, contradicting equation (5.40). Thus equation (5.42) holds.

Now suppose, in order to derive a contradiction, that equation (5.41) does not hold: there exists a state $\hat{\jmath}$ for which

$$\overline{W}_A(\hat{\jmath}) > \overline{W}_B(\hat{\jmath})\,. \tag{5.44}$$

We will construct a transition matrix for a selection process P (consistent with the fitnesses $\overline{W}_A(j)$ and $\overline{W}_B(j)$ $1 \leq j \leq N-1$) such that $p_{\text{fix}}(i) \geq i/N$, which contradicts equation (5.40) holding for all selection processes.

We restrict our attention to transition matrices such that for any $1 \leq k \leq N-1$, $P_{j,k} = 0$ if and only if $k < j-1$, $k > j+1$ or $j = k$. Thus, at any time-step and population state j, the number of individuals of type A can either increase or decrease by 1. P then defines a "birth-death" process, for which the fixation probabilities starting from state $X(0) = i$ satisfy (see appendix 5.B):

$$p_{\text{fix}}(i) = \frac{1 + \sum_{k=1}^{i-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}}}{1 + \sum_{k=1}^{N-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}}}\,. \tag{5.45}$$

Let $\mathcal{A}_+$, $\mathcal{A}_-$ and $\mathcal{A}_0$ be the sets of states at which the mean fitness of individuals of type A is higher than, lower than or equal to that of B individuals (respectively). Note that $\hat{\jmath} \in \mathcal{A}_+$ and $\hat{\imath} \in \mathcal{A}_-$. We then specify the ratios of the non-vanishing transition probabilities by

$$\frac{P_{j,j-1}}{P_{j,j+1}} = \begin{cases} r_+ & j \in \mathcal{A}_+\,, \\ r_- & j \in \mathcal{A}_-\,, \\ 1 & j \in \mathcal{A}_0\,, \end{cases} \tag{5.46}$$

with $r_- > 1$ and ,$1 > r_+ > 0$.

Observe that P is a mixed-irreducible selection process:

- If $X(t) = j$, then

$$\mathbb{E}\left(X(t+1)\right) - X(t) = (j+1)P_{j,j+1} + (j-1)P_{j,j-1} - j = P_{j,j+1} - P_{j,j-1}\,, \tag{5.47}$$

  so H1 is satisfied.

- As for the Moran process (see equation (5.15a)), for any j and k such that $1 \leq j \leq N-1$, $0 \leq k \leq N$ and $j \neq k$, there is a positive probability of transitioning from state j to state k in $d = |j-k|$ steps: setting $\sigma = \operatorname{sign}(k-j)$, we have

$$\Pr\left(X(t+d) = k | X(t) = j\right) = \prod_{m=1}^{d} P_{(j+\sigma(m-1)),(j+\sigma m)} > 0 \tag{5.48a}$$

$$\Pr\left(X(t+2) = j | X(t) = j\right) = P_{j,j+1}P_{j+1,j} + P_{j,j-1}P_{j-1,j} > 0\,, \tag{5.48b}$$

  so all states can be reached from state $X(t) = i$ in finite time. Thus, P is mixed irreducible. Moreover, the probability of B fixing at a future time $t+\tau$ ($\tau \geq 0$) is positive, satisfying H2.

- The states 0 and N are absorbing, so H3 is trivially satisfied.

For $1 \leq j \leq N-1$, we define the number of states $1 \leq k \leq j$ at which the mean fitness of A individuals is higher than that of B individuals,

$$\alpha_+(j) = \left|\left\{k | 1 \leq k \leq j \text{ and } k \in \mathcal{A}_+\right\}\right|, \tag{5.49}$$

and similarly,

$$\alpha_-(j) = \left|\left\{k | 1 \leq k \leq j \text{ and } k \in \mathcal{A}_-\right\}\right|. \tag{5.50}$$

Lastly, let $1 \leq a_+$ be the smallest number of individuals of type A in the population at which type A's mean fitness is higher than type B's, that is,

$$a_+ = \min \mathcal{A}_+ \geq 1\,. \tag{5.51}$$

Note that $a_+ \leq \hat{\jmath} < N$, and that $\alpha_+(j) = 0$ for all $j < a_+$.

From equation (5.5), the fixation probability $p_{\text{fix}}(i)$ is a rational function of r_+ and r_-,

$$p_{\text{fix}}(i) = \frac{1 + \sum_{k=1}^{i-1} r_+^{\alpha_+(k)} r_-^{\alpha_-(k)}}{1 + \sum_{k=1}^{N-1} r_+^{\alpha_+(k)} r_-^{\alpha_-(k)}} \tag{5.52}$$

and is continuous because the denominator is positive for any $r_-, r_+ > 0$.

If $i \geq a_+$, then $p_{\text{fix}}(i) \to 1$ as $r_+ \to 0$. If $i < a_+$, then

$$\lim_{r_+ \to 0} p_{\text{fix}}(i) = \frac{1 + \sum_{k=1}^{i-1} r_-^{\alpha_-(k)}}{1 + \sum_{k=1}^{a_+-1} r_-^{\alpha_-(k)}} \xrightarrow[r_- \to 1]{} \frac{i}{a_+} > i/N\,. \tag{5.53}$$

It is thus possible to choose r_- sufficiently close to 1 and r_+ sufficiently close to 0 to ensure that $p_{\text{fix}}(i) > i/N$, which completes the proof. □

5.6 Application to evolutionary game theory in finite populations

Evolutionary game theory [28, 59] is concerned with a population of agents whose fecundity (or fitness) is determined by their payoffs in interactions modelled as games. The strategies in these games are heritable traits, and the payoffs are typically dependent on which strategies other agents play. A key concept in evolutionary game theory is evolutionary stability [28, 156]. In an infinite population, the standard definition is

Definition 5.6.1 (Evolutionary stability)**.** *A strategy s is **evolutionarily stable** (ES)* iff *a single agent that plays a different strategy cannot invade the population (all strategies different from the resident strategy s are selected against).*

Typically, one says that selection opposes invasion of type B by type A if the mean fitness of a single invader of type A in a population otherwise composed of agents of type B is lower than the fitness of the agents of type B in this population (*e.g.,* [186]). Because of H1, it is possible to provide an equivalent definition in terms of the underlying selection process:

Definition 5.6.2 (Selection Opposes Invasion)**.** *For a selection process P, we say that **selection opposes invasion** of B by A if*

$$\mathbb{E}\left(X(t+1) | X(t) = 1\right) = \sum_{j=1}^{N} j P_{1,j} < 1\,, \tag{5.54}$$

and ***selection favours invasion*** *if*

$$\mathbb{E}\left(X(t+1)|X(t)=1\right)=\sum_{j=1}^{N} jP_{1,j}>1\,. \tag{5.55}$$

However, due to the inherent stochasticity of finite populations, determining whether or not selection favours invasion of mutant strategies is no longer sufficient to determine evolutionary stability in finite-population games: in a population of constant size N, if a resident strategy is invaded by a single agent playing a different strategy that is equally fit, proposition 5.5.1 implies that for any selection process, the invading strategy fixes with probability $1/N$. Moreover, the fixation probability of a strategy that is selected *against* when rare can be larger than $1/N$, if it is selected for when sufficiently common [186, 212]. Motivated by this, Nowak *et al.* have refined the definition of evolutionary stability of a strategy in a finite population to take into account the possibility of fixation of mutant strategies [186]. Their definition, though given in the context of a Moran process, applies to general selection processes:

Definition 5.6.3 (Evolutionary stability in a finite population)**.** *A strategy A is* ***evolutionarily stable*** *(ESS_N) in a population of size N* iff*, when invaded by a single mutant playing a different strategy $B \neq A$, selection opposes both invasion and fixation of B.*

- *the mutant's fitness is lower than the residents' (selection opposes invasion; definition 5.6.2),*
- *the mutant's fixation probability is less than $1/N$ (selection opposes fixation; definition 5.5.3).*

More recently, Stewart and Plotkin [213] refer to selection opposing invasion by a single mutant as "evolutionary robustness", on the grounds that the invasion dynamics are less important than which strategy fixes:

Definition 5.6.4. *A resident strategy A is* ***evolutionarily robust*** *against an invading mutant strategy B if selection opposes fixation of B (*i.e., *B's fixation probability is less than $1/N$) when a population playing A is invaded by a single mutant playing B.*

If the payoff obtained from a game with heritable strategies A and B contributes linearly to individual fitness, Lemma 5.5.4 yields intuitive conditions for evolutionary robustness and stability in finite populations: if

- the expected payoff for strategy B is no less than the expected payoff for A (at all population states to which the population can evolve from the initial one); and
- there is at least one state (to which the population can evolve from the initial state) where the expected payoff for A is less than for B;

then B is evolutionarily robust to invasion by A. If, additionally, the expected fitness of a mutant playing A in a resident population otherwise playing B is lower than the residents' expected fitness, then B is evolutionarily stable (ESS_N). We formalize these statements in Corollary 5.6.5 and explain how the assumption of linearity can be relaxed in Remark 5.6.6.

Corollary 5.6.5. *Consider a population of constant size N playing a game in which the two available strategies, A and B, are heritable traits. For any mixed-type population state i ($1 \leq i \leq N-1$), let the fitness of an agent obtaining payoff π at state i be a random variable, $W_i(\pi)$, with mean*

$$\overline{W}_i(\pi) = \mathbb{E}\left(W_i(\pi)\,|\pi\right). \tag{5.56}$$

Let the payoffs to agents playing strategy s ($s = A$ or B) be random variables, $\pi_s(i)$, with mean $\overline{\pi}_s(i)$. Denote the mean fitnesses of agents playing a strategy s at population state i by

$$\overline{W}_s(i) = \mathbb{E}\left(W_i(\pi_s(i))\right), \tag{5.57}$$

the expectation being taken over all possible payoffs to an agent playing s in a population at state i. Suppose that the following conditions hold:

(I) At any mixed-type population state, the mean payoff $\overline{\pi}_s(i)$ and the mean fitness $\overline{W}_s(i)$, of an agent playing strategy $s = A$ or B, are finite.

(II) The mean payoff of individuals of type A is never more than *type B, regardless of the number of individuals of type A in the population ($\overline{\pi}_A(i) \leq \overline{\pi}_B(i)$ for all $1 \leq i \leq N-1$).*

(III) There exists a mixed-type population state $\hat{\imath}$ accessible from the state $i = 1$ at which the mean payoff of an agent of type A is less than the mean payoff of an agent of type B ($\overline{\pi}_A(\hat{\imath}) < \overline{\pi}_B(\hat{\imath})$).

(IV) The fitness of an agent obtaining payoff π at a mixed-type state i is

$$W_i(\pi) = w_i\pi + V, \tag{5.58}$$

where $w_i > 0$ represents the strength of selection at state i and V is a real-valued random variable with finite expectation $\mathbb{E}(V) < \infty$, representing the variability in the fitness of an individual with a given payoff. We further assume that $\mathbb{E}(V)$ is independent of the payoff π, so that if an individual's payoff π is randomly distributed, its mean fitness is linear in its mean payoff,

$$\mathbb{E}\left(W_i(\pi)\right) = w_i\,\mathbb{E}(\pi) + \mathbb{E}(V), \tag{5.59}$$

(although V itself need not be independent of π).

Then strategy B is evolutionarily robust against invasion by A, for any selection process P (with respect to the frequency-dependent fitness $W_i(\pi)$, $1 \leq i \leq N-1$). If assumption (III) is satisfied for $\hat{\imath} = 1$, then strategy B is also an ESS$_N$.

Proof. From equations (5.57) and (5.59), we have

$$\overline{W}_A(i) = \mathbb{E}\left(W_i(\pi_A(i))\right) = w_i\overline{\pi}_A(i) + \mathbb{E}(V),$$

and similarly,

$$\overline{W}_B(i) = w_i\overline{\pi}_B(i) + \mathbb{E}(V).$$

Thus,

$$\overline{W}_B(i) - \overline{W}_A(i) = w_i\big(\overline{\pi}_B(i) - \overline{\pi}_A(i)\big) \geq 0, \tag{5.60}$$

with a strict inequality for $i = \hat{\imath}$. The conclusion that B is evolutionarily robust now follows immediately from lemma 5.5.4. If $\overline{\pi}_A(1) < \overline{\pi}_B(1)$, then from definition 5.6.3, B is an ESS$_\text{N}$. □

Remark 5.6.6. *If assumption (II) of corollary 5.6.5 is replaced by the stronger constraint on the game payoff distributions (rather than just their means), that for any $\phi \geq 0$,*

$$\Pr(\pi_B(i) \geq \phi) \geq \Pr(\pi_A(i) \geq \phi), \tag{5.61}$$

then assumption (IV) can be weakened to the mean fitness $\overline{W}_i(\pi)$ being some non-decreasing function of the payoff π. This follows because we have only used assumption (IV) in deducing $\overline{W}_B(i) - \overline{W}_A(i) \geq 0$ (in equation (5.60)). But if equation (5.61) holds, then since $\overline{W}_i(\pi)$ is increasing, then $\overline{W}_i(\pi_B(i)) - \overline{W}_i(\pi_A(i)) \geq 0$, so $\overline{W}_B(i) - \overline{W}_A(i) \geq 0$ still holds, with a strict inequality for $i = \hat{\imath}$.

5.7 Conclusions

We have defined a large class of biologically sensible models of selection acting on two traits in populations of N agents in the absence of mutation (definition 5.3.1), and a subclass of models of neutral drift (definition 5.3.8). Our main results are simple sufficient conditions for selection favouring or opposing fixation of a trait (lemma 5.5.4) for *any* selection process. From an entirely mathematical perspective, our analysis identifies conditions under which the inequality in the optional stopping theorem for supermartingales (theorem 5.A.6) can be made strict.

We used lemma 5.5.4 to obtain sufficient conditions for evolutionary robustness and stability in a finite population (corollary 5.6.5). In fact, lemma 5.5.6 implies that the conditions of corollary 5.6.5 *characterize* the games for which evolutionary robustness and stability are independent of the selection process . The proof of corollary 5.6.5 is simple,

but the result has important implications; in particular, it is a critical component used in chapter 4 to arrive at a complete analysis of evolutionary stability in the continuous snow-drift game for any selection process in a well-mixed population.

Focusing on fixation probabilities (as opposed to fixation times or properties of the continuum limit) allows us to maintain more generality compared to the formulation of Generalized Wright-Fisher (GWF) models [68, 75], both in removing the assumption on the second moment of the drift process [68, eq. (5)], and expanding the class of non-drift processes that are included. The importance of the latter generalization is highlighted by the fact that, as noted in [75, p. 36], the classical Wright-Fisher process with selection is *not* a GWF process, whereas (excluding mutation) it is a selection process according to definition 5.3.1 (see § 5.4.2).

Our treatment was limited to two-trait models for simplicity, but the framework can be extended to a larger number of interacting strategies in the population (at the expense of increasing the complexity of the analysis [78]). The presence of only two competing strategies in the population at any time is a common assumption in many evolutionary models: for instance, both the standard formulation of adaptive dynamics [58] and its extension to structured populations [79], rely on the assumption of "trait substitution". Under this assumption, mutants arise and either vanish or fix before a new mutation occurs. In practice, multiple mutant strategies may be present in a population at the same time if fixation rates are slow compared to mutation rates. It would therefore be useful, on the one hand, to construct a framework that relaxes the assumptions of trait substitution and, on the other hand, to identify conditions under which models based on trait substitution are valid (by comparing two-type and several-type populations subject to stronger assumptions on the selection process that allow bounds on fixation times to be obtained, *e.g.,* GWF models [68, 75]).

While we confined our analysis to asexual populations, extensions that allow for genetic inheritance in sexual populations would be useful. Such extensions, however, might depend on the particulars of the genetic system. For example, in diploid populations, the fitnesses of the two homozygotes and the heterozygote may differ. Moreover, if the allele for trait A is dominant over trait B, then populations with identical phenotypes may have vastly different genetic make-ups, which may have different transition probabilities to other states, *e.g.,* when the entire population displays the phenotype A, one cannot know how many individuals are heterozygotes. But if all individuals are homozygotic for A, then A has fixed and the transition probability to any other state is 0, which is not the case if all individuals are heterozygotes. Thus, for sexual diploid populations, the state space will likely contain information on the different genetic types in the population, rather than just the phenotypic types. Additional extensions of our framework that may prove fruitful include accounting for mutation between the two strategies, considering populations of variable size, and evolution in continuous time.

Acknowledgments

DE was supported by NSERC. CM was supported by the Ontario Trillium Foundation and the McMaster University Department of Mathematics and Statistics. We are grateful to Ben Bolker, Sarah Otto and Joshua Plotkin for valuable discussions and comments.

Appendix

5.A Definitions and theorems from probability theory

In this appendix, we collect definitions and theorems from probability theory necessary for the proofs in the main text, but that are not necessary to follow the flow of the paper when proofs are omitted. These are collected here to aid readers who are already sufficiently familiar with probability theory so as to require only a brief reminder of the relevant concepts. For a comprehensive exposition of the relevant theory, see the references cited for each definition and theorem.

All random variables are assumed to be real-valued.

5.A.1 Total expectation

Theorem 5.A.1 (Law of total expectation [80, equation 1.14])**.** *If X and Y are random variables and g is a function such that $\mathbb{E}\big(g(X)\big) < \infty$, then*

$$\mathbb{E}_X\big(g(X)\big) = \mathbb{E}_Y\big(\,\mathbb{E}_X(g(X)|Y)\,\big)\,. \tag{5.62}$$

5.A.2 Markov Chains

Let $X(t)$ be a discrete-time Markov chain with stationary transition matrix P. We say that the states i and j ***communicate*** if state i leads to state j, and *vice-versa* (see definition 5.3.3). A state that communicates with all the states it leads to is called ***essential***; otherwise, it is ***inessential***. Lastly, a state i is called ***recurrent*** if, when starting from $X(0) = i$ the process is guaranteed to return to state i (at least once), that is

$$\Pr\big(\exists \tau > 0 \text{ such that } X(\tau) = i | X(0) = i\big) = 1\,. \tag{5.63}$$

Comprehensive discussions of the classification of states of Markov chains can be found in [210, 214].

5.A.3 Martingale theory

Definition 5.A.2 (Martingales [80, chapter 6])**.** *Let $X(t)$ be a bounded, discrete-time Markov process.*

- *$X(t)$ is a martingale* iff *for all $t \geq 0$,*

$$\mathbb{E}\left(X(t+1)\right) = X(t)\,. \tag{5.64}$$

- *$X(t)$ is a supermartingale* iff *for all $t \geq 0$,*

$$\mathbb{E}\left(X(t+1)\right) \leq X(t)\,. \tag{5.65}$$

Definition 5.A.3 (Stopping time [80, chapter 6, definition 3.1])**.** *Let $X(t)$ be a discrete-time Markov process. A random variable T taking values in $\{0, 1, 2, \ldots, \infty\}$ such that the indicator function of the event $T = n$, $I_{T=n}$, is a function of $X(0), \ldots, X(n)$, is called a* ***stopping time*** *(with respect to $X(t)$).*

Thus, a random variable T is a stopping time *iff* the event "stopped at time n" ($T = n$) depends only on the states $X(t)$ that have occurred up to time n. A stopping time can be interpretted as a rule for deciding whether or not to stop a process, based only on events that have alraedy occurred (not on any knowledge of the future).

Remark 5.A.4. *Martingales, supermartingales and stopping times can be defined more generally. We use these restricted definitions for simplicity.*

The following is a version of the optional stopping theorem given in [80, chapter 6, theorem 3.1]:

Theorem 5.A.5 (Optional stopping theorem)**.** *Let X be a martingale and T a stopping time that is almost surely finite* $(\Pr(T < \infty) = 1)$ *and*

$$\mathbb{E}\left(\sup_{t\geq 0}\left|X(\min\{T,t\})\right|\right) < \infty\,. \tag{5.66}$$

Then, $\mathbb{E}\left(X(T)\right) = E\big(X(0)\big)$.

Condition (5.66) means that the expected supremum (least upper bound) of the magnitude of the state before a stopping time is finite.

Theorem 5.A.6 (Optional stopping for supermartingales [80, chapter 6, theorem 4.2])**.** *Let X be a supermartingale and T a stopping time. Then,* $\mathbb{E}\left(X(T)I_{\{T<\infty\}}\right) \leq E\big(X(0)\big)$.

5.B Fixation probabilities for birth-death processes

Suppose that individuals in a population of constant size N can possess one of two traits, A and B. Let the state of the population (*i.e.,* the number of individuals of type A) evolve according to a discrete-time ***birth-death process*** in which a trait that has disappeared cannot re-emerge. That is, the population state may change by at most one at any given time-step (individuals change their type one at a time), and the states 0 and N are absorbing. In this appendix, we find $p_{\text{fix}}(i)$, the fixation probability of the trait A, when there are initially i individuals of type A in the population. We do this following the method presented in [156].

Mathematically, the time evolution of the population composition follows a Markov process with transition matrix P satisfying

$$P_{k,k} = 1 - P_{k,k+1} - P_{k,k-1}\,, \tag{5.67}$$

and $P_{k,j} = 0$ for all $0 \leq j < k-1$ and $k+1 < j \leq N$, where $P_{k,k+1}$ and $P_{k,k-1}$ are the transition probabilities from the state in which there are k individuals of type A, to the ones in which the population contains $k+1$ or $k-1$ individuals of type A, respectively. Note also that $P_{0,0} = P_{1,1} = 1$ and $P_{0,k} = P_{N,N-k} = 0$ for all $1 \leq k \leq N$ (the states corresponding to homogeneous populations are absorbing).

Let $p_{\text{fix}}(i)$ be the probability of reaching state N (fixation of A) when starting from state i. It follows that $p_{\text{fix}}(0) = 0$, $p_{\text{fix}}(N) = 1$ and for $1 \leq i \leq N-1$,

$$p_{\text{fix}}(i) = P_{i,i-1}p_{\text{fix}}(i-1) + P_{i,i+1}p_{\text{fix}}(i+1) + P_{i,i}p_{\text{fix}}(i)\,. \tag{5.68}$$

Consequently,

$$(P_{i,i+1} + P_{i,i-1})p_{\text{fix}}(i) = (1 - P_{i,i})p_{\text{fix}}(i) = P_{i,i-1}p_{\text{fix}}(i-1) + P_{i,i+1}p_{\text{fix}}(i+1)\,,$$

so

$$P_{i,i-1}(p_{\text{fix}}(i) - p_{\text{fix}}(i-1)) = P_{i,i+1}(p_{\text{fix}}(i+1) - p_{\text{fix}}(i))\,,$$

or

$$y_{i+1} = \frac{P_{i,i-1}}{P_{i,i+1}}y_i\,,$$

where $y_i = p_{\text{fix}}(i) - p_{\text{fix}}(i-1)$ for $1 \leq i \leq N$. Thus,

$$\begin{aligned} y_1 &= p_{\text{fix}}(1) - p_{\text{fix}}(0) = p_{\text{fix}}(1) \ , \\ y_2 &= \frac{P_{1,0}}{P_{1,2}} y_1 = \frac{P_{1,0}}{P_{1,2}} p_{\text{fix}}(1) \ , \\ y_3 &= \frac{P_{2,1}}{P_{2,3}} y_2 = \frac{P_{2,1}}{P_{2,3}} \frac{P_{1,0}}{P_{1,2}} p_{\text{fix}}(1) \ , \\ &\vdots \\ y_{i+1} &= \prod_{j=1}^{i} \frac{P_{j,j-1}}{P_{j,j+1}} p_{\text{fix}}(1) \end{aligned} \tag{5.69}$$

for $2 \leq i \leq N-1$.

Summing y_k for $1 \leq k \leq i \leq N$ gives

$$\sum_{k=1}^{i} y_k = \sum_{k=1}^{i} \left(p_{\text{fix}}(k) - p_{\text{fix}}(k-1) \right) = p_{\text{fix}}(i) - p_{\text{fix}}(0) = p_{\text{fix}}(i) \ . \tag{5.70}$$

From equations (5.69) and (5.70),

$$p_{\text{fix}}(i) = y_1 + \sum_{k=1}^{i-1} y_{k+1} = p_{\text{fix}}(1) \left(1 + \sum_{k=1}^{i-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}} \right) . \tag{5.71}$$

Since $p_{\text{fix}}(N) = 1$, substituting $i = N$ in equation (5.71) gives

$$p_{\text{fix}}(1) = \frac{1}{1 + \sum_{k=1}^{N-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}}} \ . \tag{5.72}$$

Thus, from equations (5.72) and (5.71), the fixation probability of A when there are initially i individuals of type A in the population is

$$p_{\text{fix}}(i) = \frac{1 + \sum_{k=1}^{i-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}}}{1 + \sum_{k=1}^{N-1} \prod_{j=1}^{k} \frac{P_{j,j-1}}{P_{j,j+1}}} \ . \tag{5.73}$$

Chapter 6

Conclusion

This thesis began with the analysis of a specific public goods game (chapter 2), then a class of snowdrift games in infinite populations (chapter 3), general snowdrift games in finite populations (chapter 4) and, finally, general selection processes in finite populations (chapter 5). To conclude, we position this work in a broader context and comment on some possible implications.

The main question in chapter 2 on the vaccination game is essentially "What is the cost of individual free choice from the perspective of the group as a whole?" This cost stems from the existence of freeriders, who benefit from herd immunity (the public good) while avoiding the personal cost of vaccination. While we explored this issue in the context of the vaccination game, free choice may yield a mean payoff that is sub-optimal from the group perspective in other public goods games as well. Since public goods games arise in many other contexts of human interaction (*e.g.,* ozone depletion [215] or the management of forests [216] or fisheries [118]), similar analyses of the cost of free choice (*i.e.,* the lack of regulation) will likely be useful in other situations. Moreover, the existence of the mortality plateau in a general class of vaccination games suggests the existence of an analogous plateau in other instances where individual and societal interests are in conflict; the conditions for the existence of such a plaeau would likely be: (i) individuals choose between cooperation and defection, (ii) cooperation carries a constant cost (similar to the cost of vaccination in the vaccination game), and (iii) variability in the effort taken to mitigate costs at the societal level effects variability in the cost to a defecting individual (similar to the vaccination effort in the post-outbreak vaccination response).

Chapters 3 and 4 together show that in public goods games, differences in population size can result in qualitatively different evolutionary dynamics; in particular, making the approximation that a large population is actually infinite can lead to inferences that are completely different in any finite population. As stated in chapter 4, because it is common to approximate finite populations with infinite ones, the results of this thesis stress the importance of deriving clear conditions for when the infinite population approxima-

tion is valid. It seems likely that comparing the size of the interacting group with the total population size will yield a suitable criterion for the validity of the infinite population approximation, but this has yet to be shown rigorously. The predicted qualitative difference in evolutionary outcomes between finite and infinite populations results from differences in the mean number of defectors that cooperators and defectors interact with (the mean cost defectors inflict on their group members is identical in finite and infinite populations). Consequently, results will differ in populations that are not well-mixed (*e.g.,* when the games are played on a network, or when cooperators and/or defectors can identify one another).

Chapter 5 highlights the importance of martingale and Markov chain theory in analyzing selection processes in finite populations, beyond the Moran and Wright-Fisher models. We developed the theory necessary for our analysis of continuous public goods games in finite well-mixed populations (chapter 4), but did not consider processes that include mutation, which would be a natural next step. It is also worth mentioning that in the context of the existing frameworks for modelling evolution in structured populations [77, 79], it is possible to prove as a corollary of our work in chapters 4 and 5 that the structure coefficient in a well-mixed population is $\sigma = (N-2)/N$, for any updating rule (which in this case corresponds to a particular selection process, as defined in chapter 5). Lastly, we point out that our results on general selection processes may be applied even in populations that are not well-mixed. In particular, observe that theorem 4.3.1 is a particular application of the theory of general selection processes to populations that are not well-mixed. However, deriving the transition matrix for a selection process in a population that is not well-mixed is in general likely to be a computationally-intensive task (though it may be rendered more tractable if the population structure contains some symmetries, *e.g.,* a lattice structure).

Finally, we note that as seen in chapters 2, 3 and 4, the interesting phenomena observed in public goods games stem from the possibility of defectors freeloading on others' contributions. Much of the literature about the evolution of cooperation involves mechanisms by which cooperation is enforced or encouraged via rewards (*e.g.,* reciprocal altruism [139, 170, 171, 217]) or punishment [23, 218, 219]. However, other interesting approaches to encouraging cooperation have been suggested. In particular, extensive branches of the economics and game theory literatures focus on mechanism design, especially in public goods games [86, 220, 221, 222]. This field of study attempts to design rules for games, so that the outcomes are desirable in some sense (*e.g.,* socially equitable, or minimizing group cost), despite agents acting strategically. In public goods games, such mechanisms may be used to encourage cooperation by employing incentive taxes. Moreover, when agents' preferences vary (*e.g.,* as a result of variability in resources or needs), agents can in some cases be encouraged to reveal their true preference. Thus, it is sometimes possible for a central agency to obtain a socially-desirable outcome while maintaining a measure of free choice (although, personal freedom is often impinged upon by the taxation employed), in a sense resolving the group-individual conflict that was the focus of chapter 2.

While mechanism design is not without its limitations [221, §4.2.4], it would thus be

interesting to attempt to apply its tools to the vaccination game. One study makes a step in this direction [223], but assumes that the fraction of unvaccinated individuals who become ill is linear in the vaccine coverage, which is often unrealistic (see [36] and chapter 2); incorporating realistic epidemic models in order to calculate the probability that an unvaccinated individual contracts the disease may yield results that are more epidemiologically sound.

Additionally, the theory of mechanism design could in principle be applied in the context of evolutionary game theory, and specifically in relation to public goods in nature — which, to the best of our knowledge, has never been attempted. One such possibility arises in the context of eusocial insects, where it might be possible for a "queen" to act in a manner similar to central government agencies in human societies, and manufacture conditions under which workers have an incentive to cooperate (and in particular forgo reproduction). Thus, it may be possible that selection operates on the structure of the evolutionary game itself—in particular, on the costs and benefits of cooperating, as experienced by the hive workers—incentivizing cooperation, which would be a fascinating topic for future work.

Bibliography

[1] Merriam-Webster. Merriam-Webster.com;. "Cooperate". Available from: http://www.merriam-webster.com/dictionary/cooperate.

[2] Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635–1638.

[3] Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA, et al. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014;514(7520):59–64.

[4] Manser M, Brown M. Marriage and household decision-making: A bargaining analysis. International Economic Review. 1980;p. 31–44.

[5] Lundberg S, Pollak RA. Noncooperative bargaining models of marriage. The American Economic Review. 1994;p. 132–137.

[6] Yarwood R, Edwards B. Voluntary action in rural areas: The case of neighbourhood watch. Journal of Rural Studies. 1995;11(4):447–459.

[7] Weintraub S. Silent night: the story of the World War I Christmas truce. Simon and Schuster; 2001.

[8] Takahashi S. Counter A2/AD in Japan-US Defense Cooperation–Toward 'Allied Air-Sea Battle'. Washington, Project 2049. 2012;.

[9] Grieco JM. Cooperation among nations: Europe, America, and non-tariff barriers to trade. Cornell University Press; 1990.

[10] Finus M. Game Theory and International Environmental Co-operation: A Survey with an Application to the Kyoto-Protocol. Fondazione Eni Enrico Mattei; 2000.

[11] Strassmann JE, Zhu Y, Queller DC. Altruism and social cheating in the social amoeba Dictyostelium discoideum. Nature. 2000;408(6815):965–967.

[12] Ferriere R, Bronstein JL, Rinaldi S, Law R, Gauduchon M. Cheating and the evolutionary stability of mutualisms. Proceedings of the Royal Society of London B: Biological Sciences. 2002;269(1493):773–780.

[13] Wenseleers T, Helanterä H, Hart A, Ratnieks FL. Worker reproduction and policing in insect societies: an ESS analysis. Journal of Evolutionary Biology. 2004;17(5):1035–1047.

[14] Boesch C, Boesch H, Vigilant L. Cooperative hunting in chimpanzees: kinship or mutualism? In: Kappeler PM, Van Schaik CP, editors. Cooperation in primates and humans. Springer; 2006. p. 139–150.

[15] Dobzhansky T. Nothing in Biology makes sense except in the light of evolution. American Biology Teacher. 1973;35:125–129.

[16] Mayr E, Provine WB. The evolutionary synthesis: perspectives on the unification of biology. Harvard University Press; 1998.

[17] Huxley J, Pigliucci M, Müller GB. Evolution: the modern synthesis: the definitive edition. MIT Press; 2010.

[18] Dawkins R. The selfish gene. 199. Oxford University Press; 2006.

[19] Axelrod R, Hamilton WD. The evolution of cooperation. Science. 1981;211(4489):1390–1396.

[20] Motro U. Co-operation and defection: playing the field and the ESS. Journal of Theoretical Biology. 1991;151(2):145–154.

[21] Frank SA. A general model of the public goods dilemma. Journal of Evolutionary Biology. 2010;23(6):1245–1250.

[22] Boyd R, Richerson PJ. Group beneficial norms can spread rapidly in a structured population. Journal of Theoretical Biology. 2002;215(3):287–296.

[23] Fehr E, Gächter S. Altruistic punishment in humans. Nature. 2002;415(6868):137–140.

[24] Ostrom E. Governing the commons: The evolution of institutions for collective action. Cambridge University Press; 1990.

[25] Taylor P. Altruism in viscous populationsan inclusive fitness model. Evolutionary ecology. 1992;6(4):352–356.

[26] Dresher M. Games of Strategy: Theory and Applications. Prentice-Hall; 1963.

[27] Rapoport A, Chammah AM. Prisoner's dilemma: A study in conflict and cooperation. vol. 165. University of Michigan Press; 1965.

[28] Maynard Smith J. Evolution and the Theory of Games. Cambridge University Press; 1982.

[29] Doebeli M, Hauert C. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. Ecology Letters. 2005;8(7):748–766.

[30] Kümmerli R, Colliard C, Fiechter N, Petitpierre B, Russier F, Keller L. Human cooperation in social dilemmas: comparing the Snowdrift game with the Prisoner's Dilemma. Proceedings of the Royal Society of London B: Biological Sciences. 2007;274(1628):2965–2970.

[31] Dawes RM. Social dilemmas. Annual review of psychology. 1980;31(1):169–193.

[32] Bednekoff PA. Mutualism among safe, selfish sentinels: a dynamic game. The American Naturalist. 1997;150(3):373–392.

[33] Clutton-Brock TH, O'Riain M, Brotherton P, Gaynor D, Kansky R, Griffin A, et al. Selfish sentinels in cooperative mammals. Science. 1999;284(5420):1640–1644.

[34] Rainey PB, Rainey K. Evolution of cooperation and conflict in experimental bacterial populations. Nature. 2003;425(6953):72–74.

[35] Cordero OX, Ventouras LA, DeLong EF, Polz MF. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. PNAS. 2012;109(49):20059–20064.

[36] Bauch CT, Earn DJD. Vaccination and the theory of games. PNAS. 2004;101(36):13391–13394.

[37] Bauch CT, Galvani AP, Earn DJD. Group interest versus self-interest in smallpox vaccination policy. PNAS. 2003;100(18):10564–10567.

[38] Wang Z, Andrews MA, Wu ZX, Wang L, Bauch CT. Coupled disease–behavior dynamics on complex networks: A review. Physics of life reviews. 2015;15:1–29.

[39] Merrill RM. Introduction to epidemiology. Jones & Bartlett Publishers; 2013.

[40] Fine P, Eames K, Heymann DL. Herd immunity: a rough guide. Clinical Infectious Diseases. 2011;52(7):911–916.

[41] Kim TH, Johnstone J, Loeb M. Vaccine herd effect. Scandinavian Journal of Infectious Diseases. 2011;43(9):683–689.

[42] Gross PA, Hermogenes AW, Sacks HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons: a meta-analysis and review of the literature. Annals of Internal Medicine. 1995;123(7):518–527.

[43] Hanlon P, Hanlon L, Marsh V, Byass P, Shenton F, Hassan-King M, et al. Trial of an attenuated bovine rotavirus vaccine (RIT 4237) in Gambian infants. The Lancet. 1987;329(8546):1342–1345.

[44] Stiglitz JE. Economics of the public sector. WW Norton; 1988.

[45] Barrett S. Global Public Goods and International Development. In: Evans JW, Davies R, editors. Too Global To Fail: The World Bank at the Intersection of National and Global Public Policy in 2025. World Bank Publications; 2014. p. 13–18.

[46] Fenner F. Smallpox and its eradication. No. pts. 1-14 in History of international public health. World Health Organization; 1988.

[47] McNeil DG Jr. Wary of Attack With Smallpox, U.S. Buys Up a Costly Drug; 2013. Accessed: 2014-12-08. Available from: http://www.nytimes.com/2013/03/13/health/us-stockpiles-smallpox-drug-in-case-of-bioterror-attack.html.

[48] Krylova O. Predicting epidemiological transitions in infectious disease dynamics: Smallpox in historic London (1664-1930) [PhD]. McMaster University, Canada; 2011.

[49] Drescher K, Nadell CD, Stone HA, Wingreen NS, Bassler BL. Solutions to the public goods dilemma in bacterial biofilms. Current Biology. 2014;24(1):50–55.

[50] Pepper JW. Drugs that target pathogen public goods are robust against evolved drug resistance. Evolutionary Applications. 2012;5(7):757–761.

[51] Axelrod R, Axelrod DE, Pienta KJ. Evolution of cooperation among tumor cells. PNAS. 2006;103(36):13474–13479.

[52] Erwin D. A public goods approach to major evolutionary innovations. Geobiology. 2015;13(4):308–315.

[53] McInerney JO, Pisani D, Bapteste E, O' Connell MJ. The public goods hypothesis for the evolution of life on Earth. Biol Direct. 2011;6(41).

[54] Archetti M, Scheuring I. Review: Game theory of public goods in one-shot social dilemmas without assortment. Journal of Theoretical Biology. 2012;299:9–20.

[55] Doebeli M, Hauert C, Killingback T. The evolutionary origin of cooperators and defectors. Science. 2004;306(5697):859–862.

[56] Dieckmann U, Law R. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of Mathematical Biology. 1996;34(5-6):579–612.

[57] Geritz SA, Mesze G, Metz JA, et al. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology. 1998;12(1):35–57.

[58] Metz JA, Geritz SA, Meszéna G, Jacobs FJ, Van Heerwaarden JS, et al. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. Stochastic and Spatial Structures of Dynamical Systems. 1996;45:183–231.

[59] Hofbauer J, Sigmund K. Evolutionary games and population dynamics. Cambridge University Press; 1998.

[60] Ridley M. Evolution. Wiley; 2003.

[61] Hartl DL, Clark AG. Principles of Population Genetics. Sinauer Associates; 2007.

[62] Ewens WJ. Mathematical Population Genetics 1: Theoretical Introduction. vol. 27. Springer Science & Business Media; 2012.

[63] Moran PAP. The statistical processes of evolutionary theory. Clarendon Press; 1962.

[64] Wright S. Evolution in Mendelian populations. Genetics. 1931;16(2):97.

[65] Fisher RA. The genetical theory of natural selection: a complete variorum edition. Oxford University Press; 1930.

[66] Hedgecock D. Does variance in reproductive success limit effective population sizes of marine organisms. In: Beaumont AR, editor. Genetics and Evolution of Aquatic Organisms. London, U.K.: Chapman and Hall; 1994. p. 122–134.

[67] Eldon B, Wakeley J. Coalescent processes when the distribution of offspring number among individuals is highly skewed. Genetics. 2006;172(4):2621–2633.

[68] Der R, Epstein CL, Plotkin JB. Generalized population models and the nature of genetic drift. Theoretical Population Biology. 2011;80(2):80–99.

[69] Sargsyan O, Wakeley J. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. Theoretical Population Biology. 2008;74(1):104–114.

[70] Eldon B, Wakeley J. Linkage disequilibrium under skewed offspring distribution among individuals in a population. Genetics. 2008;178(3):1517–1532.

[71] Eldon B, Wakeley J. Coalescence times and FST under a skewed offspring distribution among individuals in a population. Genetics. 2009;181(2):615–629.

[72] Pitman J. Coalescents with multiple collisions. Annals of Probability. 1999;27(4):1870–1902.

[73] Sagitov S, et al. The general coalescent with asynchronous mergers of ancestral lines. Journal of Applied Probability. 1999;36(4):1116–1125.

[74] Schweinsberg J. Coalescent processes obtained from supercritical Galton–Watson processes. Stochastic Processes and their Applications. 2003;106(1):107–139.

[75] Der R. A theory of generalised population processes. Philadelphia: ProQuest; 2010.

[76] Ohtsuki H, Hauert C, Lieberman E, Nowak MA. A simple rule for the evolution of cooperation on graphs and social networks. Nature. 2006;441(7092):502–505.

[77] Tarnita CE, Ohtsuki H, Antal T, Fu F, Nowak MA. Strategy selection in structured populations. Journal of Theoretical Biology. 2009;259(3):570–581.

[78] Tarnita CE, Wage N, Nowak MA. Multiple strategies in structured populations. PNAS. 2011;108(6):2334–2337.

[79] Allen B, Nowak MA, Dieckmann U. Adaptive dynamics with interaction structure. The American Naturalist. 2013;181(6):E139–E163.

[80] Karlin S, Taylor HM. A first course in stochastic processes; 1975.

[81] Imhof LA, Nowak MA. Evolutionary game dynamics in a Wright-Fisher process. Journal of Mathematical Biology. 2006;52(5):667–681.

[82] Koplow DA. Smallpox: The Fight to Eradicate a Global Scourge. University of California Press; 2004.

[83] Christensen J. CDC: Smallpox found in NIH storage room is alive; 2014. Accessed: 2014-12-08. http://www.cnn.com/2014/07/11/health/smallpox-found-nih-alive/.

[84] Kaplan EH, Craft DL, Wein LM. Emergency response to a smallpox attack: The case for mass vaccination. PNAS. 2002;99(16):10935–10940.

[85] Anderson RM, May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford: Oxford University Press; 1991.

[86] Fudenberg D, Tirole J. Game Theory. MIT Press; 1991.

[87] Krylova O, Earn DJD. Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. Journal of the Royal Society Interface. 2013;10:20130098.

[88] (WHO) WHO. World Health Statistics 2014. Geneva: World Health Organization; 2014.

[89] Hammarlund E, Lewis MW, Hansen SG, Strelow LI, Nelson JA, Sexton GJ, et al. Duration of antiviral immunity after smallpox vaccination. Nature medicine. 2003 Sep;9(9):1131–1137. Available from: http://dx.doi.org/10.1038/nm917.

[90] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A. 1927;115:700–721.

[91] Ma J, Earn DJD. Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bulletin of Mathematical Biology*. 2006;68(3):679–702.

[92] Earn DJD, Andrews PW, Bolker BM. Population-level effects of suppressing fever. Proc R Soc Lond B. 2014;281(1778):20132570.

[93] Weisstein EW. Lambert W-Function. From MathWorld –A Wolfram Web Resource. http://mathworld.wolfram.com/LambertW-Function.html;.

[94] Corless RM, Gonnet GH, Hare DEG, Jeffrey DJ, Knuth DE. On the Lambert W function. Advances in Computational Mathematics. 1996;5(4):329–359.

[95] Eichner M, Dietz K. Transmission potential of smallpox: Estimates based on detailed data from an outbreak. American Journal of Epidemiology. 2003;158(2):110–117.

[96] Gani R, Leach S. Transmission potential of smallpox in contemporary populations. Nature. 2001;414(6865):748–751.

[97] Vink MA, Bootsma MCJ, Wallinga J. Serial Intervals of Respiratory Infectious Diseases: A Systematic Review and Analysis. American Journal of Epidemiology. 2014;180(9):865–875.

[98] Svensson A. A note on generation times in epidemic models. Mathematical Biosciences. 2007;208:300–311.

[99] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2013. Available from: http://www.R-project.org/.

[100] Soetaert K, Petzoldt T, Setzer RW. Solving Differential Equations in R: Package deSolve. Journal of Statistical Software. 2010;33(9):1–25. Available from: http://www.jstatsoft.org/v33/i09.

[101] Bishop DT, Cannings C. A generalized war of attrition. J Theor Biol. 1978 Jan;70(1):85–124.

[102] Sandholm WH. Population games and evolutionary dynamics. MIT press; 2010.

[103] Leach M, Fairhead J. Vaccine Anxieties: Global Science, Child Health and Society. Science in society. Earthscan; 2007.

[104] Brown KF, Kroll JS, Hudson MJ, Ramsay M, Green J, Long SJ, et al. Factors underlying parental decisions about combination childhood vaccinations including MMR: a systematic review. Vaccine. 2010;28(26):4235–4248.

[105] Gangarosa EJ, Galazka A, Wolfe C, Phillips L, Gangarosa R, Miller E, et al. Impact of anti-vaccine movements on pertussis control: the untold story. The Lancet. 1998;351(9099):356–361.

[106] Phillips CJ. Health economics: an introduction for health professionals. John Wiley & Sons; 2008.

[107] Hirsch MW, Smale S. Differential equations, dynamical systems, and linear algebra. Academic Press, New York; 1974.

[108] Hauert C, De Monte S, Hofbauer J, Sigmund K. Replicator dynamics for optional public good games. Journal of Theoretical Biology. 2002;218(2):187–194.

[109] Cordero OX, Ventouras LA, DeLong EF, Polz MF. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. PNAS. 2012;109(49):20059–20064.

[110] Archetti M. Evolutionary game theory of growth factor production: implications for tumour heterogeneity and resistance to therapies. British Journal of Cancer. 2013;109(4):1056–1062.

[111] Brown SP, Hochberg ME, Grenfell BT. Does multiple infection select for raised virulence? Trends in Microbiology. 2002;10(9):401–405.

[112] Brown S. Cooperation and conflict in host–manipulating parasites. Proceedings of the Royal Society of London B: Biological Sciences. 1999;266(1431):1899–1904.

[113] Leighton GM. Sex and individual differences in cooperative nest construction of sociable weavers Philetairus socius. Journal of Ornithology. 2014;155(4):927–935.

[114] Houston AI, Davies NB. The evolution of cooperation and life history in the Dunnock, Prunella modularis. In: Sibly RM, Smith RH, editors. Behavioural Ecology: Ecological consequences of adaptive behaviour. Oxford, U.K.: Blackwell Scientific Publications; 1985. p. 471–487.

[115] Savage JL, Russell AF, Johnstone RA. Maternal allocation in cooperative breeders: should mothers match or compensate for expected helper contributions? Animal Behaviour. 2015;102:189–197.

[116] Parker GA, Royle NJ, Hartley IR. Intrafamilial conflict and parental investment: a synthesis. Philosophical Transactions of the Royal Society of London B: Biological Sciences. 2002;357(1419):295–307.

[117] Reeve HK, Hölldobler B. The emergence of a superorganism through intergroup competition. PNAS. 2007;104(23):9736–9740.

[118] Kraak S. Exploring the public goods game model to overcome the Tragedy of the Commons in fisheries management. Fish and Fisheries. 2011;12(1):18–33.

[119] Browning M, Chiappori PA, Weiss Y. Economics of the Family. Cambridge University Press; 2014.

[120] Kagel JH, Roth AE. The handbook of experimental economics. Princeton University Press, Princeton, NJ; 1995.

[121] Hardin G. The tragedy of the commons. Science. 1968;162(3859):1243–1248.

[122] Dugatkin LA. Cooperation Among Animals : An Evolutionary Perspective: An Evolutionary Perspective. Oxford University Press, USA; 1997.

[123] Clutton-Brock T. Cooperation between non-kin in animal societies. Nature. 2009;462(7269):51–57.

[124] Fehr E, Gächter S. Cooperation and punishment in public goods experiments. Institute for Empirical Research in Economics working paper. 1999;(10).

[125] Gavrilets S. Collective action problem in heterogeneous groups. Philosophical Transactions of the Royal Society B. 2015 Oct;370(1683):20150016.

[126] Gokhale CS, Traulsen A. Evolutionary Multiplayer Games. Dynamic Games and Applications. 2014 Mar;4(4):468–488. Available from: http://dx.doi.org/10.1007/s13235-014-0106-2.

[127] Milinski M, Semmann D, Krambeck HJ. Reputation helps solve the tragedy of the commons. Nature. 2002;415(6870):424–426.

[128] Hauert C, Michor F, Nowak MA, Doebeli M. Synergy and discounting of cooperation in social dilemmas. Journal of Theoretical Biology. Mar;239(2):195–202.

[129] Souza MO, Pacheco JM, Santos FC. Evolution of cooperation under N-person snowdrift games. Journal of Theoretical Biology. 2009;260(4):581–588.

[130] Bach LA, Bentzen S, Alsner J, Christiansen FB. An evolutionary-game model of tumour–cell interactions: possible relevance to gene therapy. European Journal of Cancer. 2001;37(16):2116–2120.

[131] Archetti M, Ferraro DA, Christofori G. Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer. PNAS. 2015;112(6):1833–1838.

[132] Archetti M, Scheuring I. Coexistence of cooperation and defection in public goods games. Evolution. 2011;65(4):1140–1148.

[133] Archetti M. Evolutionary dynamics of the Warburg effect: glycolysis as a collective action problem among cancer cells. Journal of Theoretical Biology. 2014;341:1–8.

[134] Motro U, Cohen D. A note on vigilance behavior and stability against recognizable social parasites. Journal of Theoretical Biology. 1989;136(1):21–25.

[135] Poulin R. The evolution of parasite manipulation of host behaviour: a theoretical analysis. Parasitology. 1994;109(S1):S109–S118.

[136] McGill BJ, Brown JS. Evolutionary Game Theory and Adaptive Dynamics of Continuous Traits. Annual Review of Ecology and Systematics. 2007 Dec;38(1):403–435. Available from: http://dx.doi.org/10.1146/annurev.ecolsys.36.091704.175517.

[137] Pulliam HR, Pyke GH, Caraco T. The scanning behavior of juncos: a game-theoretical approach. Journal of Theoretical Biology. 1982;95(1):89–103.

[138] Cornforth DM, Sumpter DJ, Brown SP, Brännström Å. Synergy and group size in microbial cooperation. The American Naturalist. 2012;180(3):296.

[139] Killingback T, Doebeli M. The continuous prisoners dilemma and the evolution of cooperation through reciprocal altruism with variable investment. The American Naturalist. 2002;160(4):421–438.

[140] Fujita H, Aoki S, Kawaguchi M. Evolutionary Dynamics of Nitrogen Fixation in the Legume–Rhizobia Symbiosis. PLoS ONE. 2014 Apr;9(4):e93670. Available from: http://dx.doi.org/10.1371/journal.pone.0093670.

[141] Brown JS, Vincent TL. Evolution of cooperation with shared costs and benefits. Proceedings of the Royal Society B. 2008 Sep;275(1646):1985–1994. Available from: http://dx.doi.org/10.1098/rspb.2007.1685.

[142] Santos FC, Santos MD, Pacheco JM. Social diversity promotes the emergence of cooperation in public goods games. Nature. 2008 Jul;454(7201):213–216. Available from: http://dx.doi.org/10.1038/nature06940.

[143] Hauert C. Cooperation, collectives formation and specialization. Advances in Complex Systems. 2006;9(04):315–335.

[144] Hauert C. Evolutionary dynamics. In: Skjeltorp AT, Belushkin AV, editors. Evolution from Cellular to Social Scales. Springer; 2008. p. 11–44.

[145] Liang H, Cao M, Wang X. Analysis and shifting of stochastically stable equilibria for evolutionary snowdrift games. Systems & Control Letters. 2015 Nov;85:16–22. Available from: http://dx.doi.org/10.1016/j.sysconle.2015.08.004.

[146] Ito K, Ohtsuki H, Yamauchi A. Relationship between aggregation of rewards and the possibility of polymorphism in continuous snowdrift games. Journal of Theoretical Biology. 2015 May;372:47–53. Available from: http://dx.doi.org/10.1016/j.jtbi.2015.02.015.

[147] Sasaki T, Okada I. Cheating is evolutionarily assimilated with cooperation in the continuous snowdrift game. Biosystems. 2015 May;131:51–59. Available from: http://dx.doi.org/10.1016/j.biosystems.2015.04.002.

[148] Brännström Å, Gross T, Blasius B, Dieckmann U. Consequences of fluctuating group size for the evolution of cooperation. Journal of Mathematical Biology. 2010 Oct;63(2):263–281. Available from: http://dx.doi.org/10.1007/s00285-010-0367-3.

[149] Hauert C, Holmes M, Doebeli M. Evolutionary games and population dynamics: maintenance of cooperation in public goods games. Proceedings of the Royal Society B. 2006 Oct;273(1600):2565–2571. Available from: http://dx.doi.org/10.1098/rspb.2006.3600.

[150] Wu T, Fu F, Wang L. Partner selections in public goods games with constant group

size. Physical Review E. 2009 Aug;80(2). Available from: http://dx.doi.org/10.1103/PhysRevE.80.026121.

[151] Deng K, Chu T. Adaptive Evolution of Cooperation through Darwinian Dynamics in Public Goods Games. PLoS ONE. 2011 Oct;6(10):e25496. Available from: http://dx.doi.org/10.1371/journal.pone.0025496.

[152] Maynard Smith J, Price G. lhe Logic of Animal Conflict. Nature. 1973;246:15.

[153] Chen X, Szolnoki A, Perc M, Wang L. Impact of generalized benefit functions on the evolution of cooperation in spatial public goods games with continuous strategies. Physical Review E. 2012 Jun;85(6). Available from: http://dx.doi.org/10.1103/PhysRevE.85.066133.

[154] Box GE. Science and statistics. Journal of the American Statistical Association. 1976;71(356):791–799.

[155] Killingback T, Doebeli M, Hauert C. Diversity of cooperation in the tragedy of the commons. Biological Theory. 2010;5:3–6.

[156] Nowak MA. Evolutionary dynamics: Exploring the equations of life. Harvard University Press; 2006.

[157] Taylor PD. Evolutionary stability in one-parameter models under weak selection. Theoretical Population Biology. 1989;36(2):125–143.

[158] Christiansen FB. On conditions for evolutionary stability for a continuously varying character. American Naturalist. 1991;p. 37–50.

[159] Nowak MA, Sigmund K. Evolutionary dynamics of biological games. Science. 2004;303(5659):793–799.

[160] Krupp D, Taylor PD. Social evolution in the shadow of asymmetrical relatedness. Proceedings of the Royal Society of London B: Biological Sciences. 2015;282(1807):20150142.

[161] Gardner A, West S. Demography, altruism, and the benefits of budding. Journal of Evolutionary Biology. 2006;19(5):1707–1716.

[162] Kümmerli R, Gardner A, West SA, Griffin AS. Limited dispersal, budding dispersal, and cooperation: an experimental study. Evolution. 2009;63(4):939–949.

[163] Clutton-Brock T. Breeding together: kin selection and mutualism in cooperative vertebrates. Science. 2002;296(5565):69–72.

[164] Pfeiffer T, Bonhoeffer S. An evolutionary scenario for the transition to undifferentiated multicellularity. Proceedings of the National Academy of Sciences. 2003;100(3):1095–1098.

[165] Buma B, Wessman C. Disturbance interactions can impact resilience mechanisms of forests. Ecosphere. 2011;2(5):art64.

[166] Dale VH, Joyce LA, McNulty S, Neilson RP, Ayres MP, Flannigan MD, et al. Climate change and forest disturbances: climate change can affect forests by altering the frequency, intensity, duration, and timing of fire, drought, introduced species, insect and pathogen outbreaks, hurricanes, windstorms, ice storms, or landslides. BioScience. 2001;51(9):723–734.

[167] Ross SM. A First Course in Probability. Pearson Prentice Hall; 2010.

[168] Barker JL, Barclay P, Reeve HK. Within-group competition reduces cooperation and payoffs in human groups. Behavioral Ecology. 2012;23(4):735–741.

[169] Krupp DB, Debruine LM, Barclay P. A cue of kinship promotes cooperation for the public good. Evolution and Human Behavior. 2008;29(1):49–55.

[170] Nowak MA, Sigmund K. Evolution of indirect reciprocity by image scoring. Nature. 1998;393(6685):573–577.

[171] Trivers RL. The evolution of reciprocal altruism. Quarterly Review of Biology. 1971;p. 35–57.

[172] McNamara JM, Barta Z, Fromhage L, Houston AI. The coevolution of choosiness and cooperation. Nature. 2008 Jan;451(7175):189–192. Available from: http://dx.doi.org/10.1038/nature06455.

[173] Hauert C, Doebeli M. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature. 2004;428(6983):643–646.

[174] Lotem A, Fishman MA, Stone L. Evolution of cooperation between individuals. Nature. 1999;400(6741):226–227.

[175] McNamara JM, Gasson CE, Houston AI. Incorporating rules for responding into evolutionary games. Nature. 1999;401(6751):368–371.

[176] Wright J, Cuthill I. Biparental care: short-term manipulation of partner contribution and brood size in the starling, Sturnus vulgaris. Behavioral Ecology. 1990;1(2):116–124.

[177] Markman S, Yom-Tov Y, Wright J. Male parental care in the orange-tufted sunbird: behavioural adjustments in provisioning and nest guarding effort. Animal Behaviour. 1995;50(3):655–669.

[178] Hirsch MW, Smale S, Devaney RL. Differential equations, dynamical systems, and an introduction to chaos. 3rd ed. Waltham, MA: Academic Press; 2013.

[179] Darwin C. On the origin of the species by natural selection. 1859;.

[180] Freeman S, Herron JC, HHodin JA, Miner B, Sidor C. Evolutionary analysis. Pearson Prentice Hall Upper Saddle River, NJ; 2007.

[181] Wilson EO. Sociobiology. Harvard University Press; 2000.

[182] Varian HR. Microeconomic analysis. WW Norton; 1992.

[183] Marschall MJ. Citizen participation and the neighborhood context: A new look at the coproduction of local public goods. Political Research Quarterly. 2004;57(2):231–244.

[184] Lehmann L. The stationary distribution of a continuously varying strategy in a class-structured population under mutation–selection–drift balance. Journal of evolutionary biology. 2012;25(4):770–787.

[185] Zheng DF, Yin H, Chan CH, Hui P. Cooperative behavior in a model of evolutionary snowdrift games with N-person interactions. EPL (Europhysics Letters). 2007;80(1):18002.

[186] Nowak MA, Sasaki A, Taylor C, Fudenberg D. Emergence of cooperation and evolutionary stability in finite populations. Nature. 2004;428(6983):646–650.

[187] Crawford VP. Nash equilibrium and evolutionary stability in large-and finite-population playing the field models. Journal of Theoretical Biology. 1990;145(1):83–94.

[188] Wakano JY, Iwasa Y. Evolutionary branching in a finite population: deterministic branching versus stochastic branching. Genetics. 2012;p. 229–241.

[189] Hauert C, Traulsen A, née Brandt HDS, Nowak MA, Sigmund K. Public goods with punishment and abstaining in finite and infinite populations. Biological Theory. 2008;3(2):114–122.

[190] Fogel GB, Andrews PC, Fogel DB. On the instability of evolutionary stable strategies in small populations. Ecological Modelling. 1998;109(3):283–294.

[191] Antal T, Nowak MA, Traulsen A. Strategy abundance in $2\times$ 2 games for arbitrary mutation rates. Journal of Theoretical Biology. 2009;257(2):340–344.

[192] Traulsen A, Claussen JC, Hauert C. Coevolutionary dynamics in large, but finite populations. Physical Review E. 2006;74(1):011901.

[193] Lessard S. Long-term stability from fixation probabilities in finite populations: new perspectives for ESS theory. Theoretical population Biology. 2005;68(1):19–27.

[194] Wakano JY, Lehmann L. Evolutionary and convergence stability for continuous phenotypes in finite populations derived from two-allele models. Journal of Theoretical Biology. 2012;310:206–215.

[195] Li K, Cong R, Wu T, Wang L. Social exclusion in finite populations. Physical Review E. 2015 Apr;91(4). Available from: http://dx.doi.org/10.1103/PhysRevE.91.042810.

[196] Zhong LX, Qiu T, Xu JR. Heterogeneity Improves Cooperation in Continuous Snowdrift Game. Chinese Physics Letters. 2008 May;25(6):2315–2318. Available from: http://dx.doi.org/10.1088/0256-307X/25/6/107.

[197] Zhang Y, Fu F, Wu T, Xie G, Wang L. A tale of two contribution mechanisms for nonlinear public goods. Scientific Reports. 2013 Jun;3. Available from: http://dx.doi.org/10.1038/srep02021.

[198] Zhang Y, Wu T, Chen X, Xie G, Wang L. Mixed strategy under generalized public goods games. Journal of Theoretical Biology. 2013 Oct;334:52–60. Available from: http://dx.doi.org/10.1016/j.jtbi.2013.05.011.

[199] Doebeli M. Adaptive dynamics: a framework for modeling the long-term evolutionary dynamics of quantitative traits. In: Svensson E, Calsbeek R, editors. The adaptive landscape in evolutionary biology. Oxford, U.K.: Oxford University Press; 2012. p. 227–242.

[200] Waxman D, Gavrilets S. 20 questions on adaptive dynamics. Journal of Evolutionary Biology. 2005;18(5):1139–1154.

[201] Geritz SA. Resident-invader dynamics and the coexistence of similar strategies. Journal of Mathematical Biology. 2005;50(1):67–82.

[202] Cannings C. The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. Advances in Applied Probability. 1974;p. 260–290.

[203] Chia A, Watterson G. Demographic Effects on the Rate of Genetic Evolution: I. Constant Size Populations with Two Genotypes. Journal of Applied Probability. 1969;p. 231–248.

[204] Karlin S, McGregor J. Direct product branching processes and related Markov chains. PNAS. 1964;51(4):598.

[205] Huillet T, Möhle M. Population genetics models with skewed fertilities: a forward and backward analysis. Stochastic Models. 2011;27(3):521–554.

[206] Der R, Epstein C, Plotkin JB. Dynamics of neutral and selected alleles when the offspring distribution is skewed. Genetics. 2012;191(4):1331–1344.

[207] Ohtsuki H. Stochastic evolutionary dynamics of bimatrix games. Journal of Theoretical Biology. 2010;264(1):136–142.

[208] Kurokawa S, Ihara Y. Emergence of cooperation in public goods games. Proceedings of the Royal Society of London B: Biological Sciences. 2009;276(1660):1379–1384.

[209] Wild G, Taylor PD. Fitness and evolutionary stability in game theoretic models of finite populations. Proceedings of the Royal Society of London B: Biological Sciences. 2004;271(1555):2345–2349.

[210] Chung KL. Markov Chains: With Stationary Transition Probabilities. Springer-Verlag; 1967.

[211] Proulx SR. The ESS under spatial variation with applications to sex allocation. Theoretical Population Biology. 2000;58(1):33–47.

[212] Proulx S, Day T. What can invasion analyses tell us about evolution under stochasticity in finite populations? Selection. 2002;2(1-2):2–15.

[213] Stewart AJ, Plotkin JB. From extortion to generosity, evolution in the iterated prisoners dilemma. PNAS. 2013;110(38):15348–15353.

[214] Seneta E. Non-negative matrices and Markov chains. Springer Science & Business Media; 2006.

[215] Murdoch JC, Sandler T. The voluntary provision of a pure public good: The case of reduced CFC emissions and the Montreal Protocol. Journal of Public Economics. 1997;63(3):331–349.

[216] Merlo M, Briales ER. Public goods and externalities linked to Mediterranean forests: economic nature and policy. Land use policy. 2000;17(3):197–208.

[217] Fehr E, Schmidt KM. The economics of fairness, reciprocity and altruism–experimental evidence and new theories. In: Kolm SC, Ythier JM, editors. Volume 1: Foundations. Handbook of the economics of Giving, Altruism and Reciprocity. Amsterdam: Elsevier; 2006. p. 615–691.

[218] Boyd R, Richerson PJ. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology and Sociobiology. 1992;13(3):171–195.

[219] Barclay P. Reputational benefits for altruistic punishment. Evolution and Human Behavior. 2006;27(5):325–344.

[220] Shoham Y, Leyton-Brown K. Multiagent systems: Algorithmic, game-theoretic, and logical foundations. Cambridge University Press; 2008.

[221] Batina RG, Ihori T. Public goods: theories and evidence. Springer Science & Business Media; 2005.

[222] Hurwicz L, Reiter S. Designing economic mechanisms. Cambridge University Press; 2006.

[223] Manski CF. Vaccination with partial knowledge of external effectiveness. PNAS. 2010;107(9):3953–3960.