

Evolutionary Game Theory and the Adaptive  
Dynamics Approach: Adaptation where  
Individuals Interact

Piret Avila and Charles Mullon



# Evolutionary Game Theory and the Adaptive Dynamics

## Approach: Adaptation where Individuals Interact

Piret Avila<sup>1</sup> and Charles Mullon<sup>2,\*</sup>

<sup>1</sup>Institute for Advanced Studies in Toulouse, Université Toulouse 1 Capitole, 31080 Toulouse, France

<sup>2</sup>Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

\*Correspondence: [charles.mullon@unil.ch](mailto:charles.mullon@unil.ch)

**Abstract.** Evolutionary game theory and the adaptive dynamics approach have made invaluable contributions to understand how gradual evolution leads to adaptation when individuals interact. Here, we review some of the basic tools that have come out of these contributions to model the evolution of quantitative traits in complex populations. We collect together mathematical expressions that describe directional and disruptive selection in class- and group-structured populations in terms of individual fitness, with the aims of bridging different models and interpreting selection. In particular, our review of disruptive selection suggests there are two main paths that can lead to diversity: (i) when individual fitness increases more than linearly with trait expression; (ii) when trait expression simultaneously increases the probability that an individual is in a certain context (e.g. a given age, sex, habitat, size or social environment) and fitness in that context. We provide various examples of these and more broadly argue that population structure lays the ground for the emergence of polymorphism with unique characteristics. Beyond this, we hope that the descriptions of selection we present here help see the tight links among fundamental branches of evolutionary biology, from life-history to social evolution through evolutionary ecology, and thus favour further their integration.

## 1 Introduction

Owing to exponential growth, a population whose members survive and reproduce independently from one another either rapidly goes extinct, or eventually fills the universe [1–3]. Our world, however, is not limitless. Individuals therefore inevitably interact with one another, either directly, such as through fights to control breeding territories or cooperative behaviours to exploit the environment more efficiently, or indirectly, as via

the consumption of a common resource or the sharing of a common predator. When individuals vary in heritable traits that influence such interactions, those traits that are better adapted, in other words those that are associated with greater reproductive success, become more common. This sets in motion Darwinian evolution and ultimately leads to the apparent “fitness of form and function” [4] that characterizes the natural world.

While it should be easy to conceive how natural selection shapes organisms to become adapted to an environment that is determined by exogenous factors (e.g. level of precipitation, pressure, or temperature), it is less straightforward when the environment is dynamic and endogenous, made up of conspecifics whose evolving traits influence an individual’s reproductive success. Understanding adaptation when reproduction and survival depend on interactions among individuals has been one of the driving goals of evolutionary game theory and adaptive dynamics [5–9]. Such dependency on interactions can be organised in two categories: (i) an individual’s ability to reproduce may be influenced by population density, in particular reproduction must eventually be curtailed by density (density-dependence); (ii) the reproductive success of an individual depends on the traits expressed by other individuals in the population and on their frequency (frequency-dependence, [10] for further considerations). Traditional evolutionary game theory has mainly focused on understanding the consequences of frequency-dependence through direct interactions among individuals, such as helping or harming [11, 12, for textbooks]. The adaptive dynamics approach grew out of evolutionary game theory to focus mostly on the gradual evolution of quantitative traits (so when traits are subject to rare mutations with small phenotypic effects) that experience both density- and frequency-dependent selection, where frequency-dependence is typically due to indirect interactions mediated by the environment or ecology, such as apparent competition [13–15, for textbooks].

Irrespective of the specific point of focus, evolutionary game theory and the adaptive dynamics approach agree on how to characterize adaptation: a population that is adapted should be uninvadable, i.e. be resistant to invasion by any rare mutant strategy [7–9, 16–20]. This is made formal using what is commonly referred to as invasion fitness, which in a population reproducing and censused at discrete time points, is the geometric growth rate of a rare mutant coding for a strategy alternative to those in the resident population [21–25]. A strategy that maximises invasion fitness when the resident population is at the uninvadable state can be considered as optimal: it is an end-point where evolution comes to a halt [7, 16–19]. Building on evolutionary game theory, the approach of adaptive dynamics has laid a framework based on invasion fitness to understand how a population may become uninvadable under gradual evolution. In particular, this approach determines whether an uninvadable population consists of individuals all expressing the same strategy (i.e. is monomorphic) or, owing to frequency-dependence and disruptive selection, consists of multiple coexisting strategies (i.e. is polymorphic; the set of such coexisting strategies is sometimes referred to as an evolutionarily stable [4] or steady [3] coalition). The framework of adaptive dynamics is thus particularly useful to investigate the conditions that favour the emergence of phenotypic variation in the form of adaptive polymorphism, and in the context of evolutionary games, of variation in social behaviour.

More broadly, analyses of selection based on invasion fitness have made the study of adaptation possible under

an extraordinary wide range of scenarios where frequency- and density-dependence arise (e.g. competition for resources or mates, sex allocation, altruism, warfare, state-dependent life-histories, phenotypic plasticity, social learning [11, 12, 26–31, for overviews]). While the range of social and ecological scenarios afforded by this approach may seem limitless, using invasion fitness from first principles is not always straightforward owing to two reasons [32]. The first is conceptual. Loosely speaking, invasion fitness is a gene- or replicator-centered measure of reproductive success [33, 34]. But the fundamental unit of behavioral and evolutionary ecology is typically the individual organism [11, 35]. It is therefore desirable to understand adaptation at this level. A second issue is computational. In heterogeneous populations (e.g. where individuals vary in age or size and this variation influences the effects of traits on reproduction), invasion fitness turns out to be the dominant eigenvalue of a matrix which can be large and complicated [36–41]. Analyzing this eigenvalue mathematically is often cumbersome, limiting analysis and obscuring biological interpretation.

Here we review key equations that resolve some of these issues. Our goal is two-fold: to facilitate the interpretation of natural selection in terms of the individual; and to collect together simple formulas to investigate gradual evolution in class- and group-structured populations (e.g. where individuals vary in age, sex, or physiology and interact in small social groups). In particular, we provide expressions for directional and disruptive selection that are sufficient to determine whether gradual evolution leads a population to a monomorphic uninvadable state or to become polymorphic under the adaptive dynamics framework. Most of the formulas presented here are currently disseminated in the literature (sometimes derived via other frameworks, such as population or quantitative genetics), some we re-derive using invasion analysis and also extend (in our Appendix). Ultimately, we aim to provide a point of entry for other evolutionary biologists interested in modelling Darwinian evolution in non-homogeneous populations, where social and ecological interactions lead to density- and frequency-dependence and possibly adaptive polymorphism.

## 2 The basics

First, we go over the basics of invasion analysis and how it can be used to model gradual evolution within the framework of adaptive dynamics.

### 2.1 Invasion fitness and uninvadability

Suppose we are interested in the evolution of a quantitative trait, such as the extraction rate on a resource, the investment into parental care, or the proclivity to disperse. We focus our attention to scalar-valued traits for now, but later consider traits whose expression can change over ontogeny and plastic traits (see also [42–46] for the joint evolution of multiple traits). We first assume that the individuals of a large asexual population of haploid individuals all express the same value  $z \in Z$  for this trait (where  $Z \subseteq \mathbb{R}$  is the space of all strategies; Table I for a list of key symbols). If this monomorphic resident population experiences ecological or demographic changes, we wait enough time for these changes to reach a stationary state, e.g. for the population to reach an

equilibrium size or for its resource to reach a stable density (this stationary state may be more complicated like a limit cycle if there are deterministic fluctuations; or a probability distribution if there are stochastic effects). Such “ecological” attractor we can denote by  $\hat{n}(z)$  to highlight that it may depend on the resident trait. Against this background, we introduce a single copy mutation that causes the expression of an alternative trait value  $z_m \in Z$  (with subscript “m” for mutant). Over time, this copy may create a lineage whose members reproduce such that on average, they more than replace themselves. In this case, the mutant has a chance to invade and fix, which would lead to trait evolution and in turn potential ecological changes (as  $z_m$  substitutes  $z$  and ecology changes to  $\hat{n}(z_m)$ ).

To formalize the above, we define the invasion fitness  $\rho(z_m, z)$  of a mutant  $z_m$  in a resident population  $z$  as its geometric growth rate, i.e. as the per-capita per-time-step number of mutant copies produced by the mutant lineage ([21–23, 25, 47–49]; as the mutant is rare, its ecological background is set by the resident so  $\rho(z_m, z)$  will also typically depend on  $\hat{n}(z)$  but we do not write such dependency for ease of presentation). From the theory of branching processes [50], it follows that the mutant goes extinct with probability one if, and only if,  $\rho(z_m, z) \leq 1$ , i.e. if on average the mutant at most replaces itself. Otherwise, there is a non-zero probability that the mutant persists indefinitely. The definition of uninvasibility derives naturally from this result: a population monomorphic for  $z_u$  is said to be uninvasible when

$$\rho(z_m, z_u) \leq 1 \quad \text{for all } z_m \in Z, \quad (1)$$

such that it is protected against invasion from all possible mutants (e.g. [16–19])<sup>1</sup>.

## 2.2 Local analysis and gradual evolution

A related but different question is whether a population can become uninvasible through gradual evolution, to which the adaptive dynamics approach provides an answer. The main assumption behind this approach is that mutations are rare, so that a population monomorphic for  $z$  has time to reach its ecological equilibrium  $\hat{n}(z)$  before a mutant appears. Assuming further that mutations have weak unbiased phenotypic effects (i.e. following the continuum of allele model [51–54] with  $e = z_m - z$  small), invasion fitness can be Taylor expanded in  $z_m$  around  $z_m = z$  as

$$\rho(z_m, z) = 1 + s(z)(z_m - z) + \frac{1}{2}h(z)(z_m - z)^2 + O(\epsilon^3), \quad (2)$$

where

$$s(z) = \left. \frac{\partial \rho(z_m, z)}{\partial z_m} \right|_{z_m=z} \quad \text{and} \quad h(z) = \left. \frac{\partial^2 \rho(z_m, z)}{\partial z_m^2} \right|_{z_m=z} \quad (3)$$

are two key functions that respectively capture directional and disruptive selection, which can be used to characterise gradual evolution.

<sup>1</sup>Using the fact that a neutral mutant has invasion fitness equal to one (i.e. that  $\rho(z, z) = 1$ ), uninvasibility eq. (1) of a population monomorphic for  $z_u$  can equivalently be expressed as  $z_u = \arg \max_{z_m \in Z} \rho(z_m, z_u)$ , i.e.  $z_u$  maximises invasion fitness when the resident is at the uninvasible state.

The function  $s(z)$ , which is sometimes referred to as the selection gradient or local fitness gradient [9], gives the direction of selection. It tells us that selection favours mutants that increase the trait value ( $z_m > z$ ) when  $s(z) > 0$ , and conversely mutants that decrease the trait value ( $z_m < z$ ) when  $s(z) < 0$ . A trait value  $z^*$  is called a singular strategy when it is such that when expressed by the whole population, there is no directional selection, i.e.

$$z^* \text{ is a singular strategy if } s(z^*) = 0. \quad (4)$$

When the population is away from a singular strategy ( $z \neq z^*$ ),  $s(z)$  is sufficient to determine whether a nearby mutant goes extinct with certainty or whether it has a chance to invade (as we can ignore terms of order  $(z_m - z)^2$  and above in eq. 2 when  $z_m - z$  is small). If such a mutant invades, then the mutant eventually fixes and thus replaces the resident. This “invasion implies substitution” result, which has been proven to hold under a wide range of situations [27, 55–60], means that there exists a regime of rare mutation with weak effects such that evolution proceeds by a trait substitution sequence whereby the population can be thought of as transiting from one monomorphic state to another [3]. A singular strategy  $z^*$  is then approached gradually via such a sequence when

$$s'(z^*) = \left. \frac{ds(z)}{dz} \right|_{z=z^*} = h(z^*) + \left. \frac{\partial^2 \rho(z_m, z)}{\partial z_m \partial z} \right|_{z_m=z=z^*} < 0 \quad (5)$$

[61–64]. A singular strategy  $z^*$  satisfying eq. (5) is thus an attractor of selection and said to be convergence stable.

Once the population has evolved to express a singular strategy  $z^*$ , selection is determined by  $h(z^*)$  (see eq. 2 with  $s(z^*) = 0$ ). In particular, all nearby mutants are counter-selected when

$$h(z^*) < 0. \quad (6)$$

In this case, the population is locally uninvadable and selection on the trait is stabilising for  $z^*$  (Fig. 1A; [7, 62]). By contrast, any mutant can invade when  $h(z^*) > 0$ . In fact, when  $h(z^*) > 0$  and the singular strategy  $z^*$  is convergence stable (eq. 5 holds), selection is frequency-dependent and disruptive such that the population becomes polymorphic through a process referred to as evolutionary branching whereby the population trait distribution goes from unimodal to bimodal (Fig. 1B; [8, 9, 14, 27, 65, 66]). Evolutionary branching, which has been found to occur under a wide range of ecological and social interactions (like mutualism, helping, or competition [28]), can help understand how adaptive polymorphism gradually emerges in populations that are initially monomorphic.

Together, the functions  $s(z)$  and  $h(z)$  thus determine whether gradual evolution leads a population to a monomorphic uninvadable state under stabilising selection or to become polymorphic due to disruptive selection (Fig. 1). There exist techniques based on invasion analysis to model the long-term fate of adaptive polymorphism (e.g. how many morphs eventually coexist, their trait values and frequencies [9]) but here we focus on whether polymorphism emerges or not.

### 2.3 All is well in a well-mixed and homogeneous world

One fundamental aspect of  $s(z)$  and  $h(z)$  to keep in mind is that they are defined from a mutant's invasion fitness (eq. 3), which is a measure of reproductive success at the level of the gene or the replicator that causes the expression of the mutant phenotype. On conceptual and empirical grounds, it is however often desirable to understand selection at the level of the individual. In populations that are homogeneous and well-mixed (i.e. where variation in reproductive success is only associated to variation at the locus underlying the evolving trait and where individuals interact and compete at random such that a rare mutant only ever interacts with residents), this is not a problem as invasion fitness is equal to individual fitness. To see this, let  $N_t$  be the number of mutant individuals at some demographic time point  $t$ . Invasion fitness is defined by the dynamical equation  $N_{t+1} = \rho(z_m, z)N_t$  (as invasion fitness is the geometric growth rate). But this dynamical equation can equivalently be written as

$$N_{t+1} = w(z_m, z)N_t, \quad (7)$$

where  $w(z_m, z)$  is *individual* fitness: the expected number of direct descendants produced between two time points by a mutant with trait value  $z_m$ , when the resident population expresses  $z$ . It is typically only a matter of bookkeeping to piece together such an individual fitness function  $w(z_m, z)$  for a given scenario, including where interactions are frequency- and density-dependent. This in turn allows us to straightforwardly proceed with the analysis described in section 2.2 and gain insights into the outcome of evolutionary dynamics, in particular whether disruptive selection leads to the emergence of polymorphism (see Appendix A.1 for a worked out example).

The equivalence between invasion and individual fitness means that in an uninhabitable population, the expressed strategy maximises individual fitness (against itself, as in eq. 1). This offers a clear view on adaptation in well-mixed and homogeneous populations: selection leads to the expression of genetic traits that maximise the reproductive success of its bearer (in the absence of genetic conflicts). In fact, if individual fitness increases with a single intermediate quantity, such as fecundity, or in the context of evolutionary games, material payoff resulting from social interactions, then any  $z^*$  that maximises payoff also maximises individual fitness, having

$$s(z) = \left. \frac{\partial w(z_m, z)}{\partial z_m} \right|_{z_m=z} \propto \left. \frac{\partial \pi(z_m, z)}{\partial z_m} \right|_{z_m=z}, \quad \text{and} \quad h(z) = \left. \frac{\partial^2 w(z_m, z)}{\partial z_m^2} \right|_{z_m=z} \propto \left. \frac{\partial^2 \pi(z_m, z)}{\partial z_m^2} \right|_{z_m=z}, \quad (8)$$

where  $\pi(z_m, z)$  is the payoff obtained by a mutant individual in a resident population (Appendix A.2 for details). This is the basis of many optimization models in evolutionary ecology, which rather than seeking maximums of individual fitness, seek maximums of such fitness proxies, like fecundity or prey caught or rate of calorie intake [e.g. 67, 68]. Eq. (8) is also useful from an empirical point of view. As individual fitness is hard to estimate in natural populations, assays typically turn to fitness proxies, such as body weight, number of females mated, clutch size or number of seeds produced. According to eq. (8), information on the nature of selection can be yielded by performing a quadratic regression of such proxies on individual trait values (with directional selection given by the linear term and disruptive selection by the quadratic term of the regression [69, 70]).

Notwithstanding remaining technical challenges for adaptive dynamics in homogeneous and well-mixed populations [71], the equivalence between fitness at the level of the gene and of the individual thus yields many insights into trait evolution and its resulting ecological or demographic transformations (Appendix A.3 for an example of such insight). Most natural populations, however, are not homogeneous or well-mixed. In the next two sections, we review how invasion analyses can be performed for more complex populations. In particular, we provide expressions for directional ( $s(z)$ ) and disruptive selection ( $h(z^*)$ ) in terms of individual fitness to be compared with eq. (8). This aims to facilitate cross-talk among models with individual fitness as common vocabulary, and ultimately aid with the interpretation of adaptation.

### 3 Class-structured populations: reproductive values as the exchange rate between fitness components

Populations often show heterogeneity among individuals: there can be males and females, individuals of different ages or stages, individuals in different physiological states, or individuals living in different habitats. This is more generally referred to as class-structure and such structure is relevant for evolution when the fitness effect of evolving traits depends on the class of the individual expressing it. In this section, we collect together existing equations as well as present some new ones for understanding selection under three common types of class-structure (but where the population is still well-mixed such that a rare mutant only ever interacts with residents). We assume that trait expression is fixed across classes so that there is no class-specificity in trait expression (though see Box I). Technical details can be found in Appendix B.

#### 3.1 Matrix population models

We first consider a population that is divided into a finite number  $M$  of discrete classes (e.g. males and females, juvenile and mature individuals, or subordinate and dominant individuals) such that mutant dynamics can be modelled by a matrix equation,

$$N_{t+1} = \mathbf{W}(z_m, z) \cdot N_t \quad (9)$$

where entry  $i \in \{1, \dots, M\}$  of the vector  $N_t$  gives the number of mutants in class  $i$  at some time  $t$ , and the  $(i, j)$ -entry of the  $M \times M$  matrix,  $\mathbf{W}(z_m, z)$ , which we denote by  $w_{ij}(z_m, z)$ , is the expected number of mutants in class  $i$  produced by a mutant in class  $j$  between two time points (Table II for key symbols used in sections 3 and 4). The matrix  $\mathbf{W}(z_m, z)$  is sometimes referred to as the mean matrix in the theory of multi-type branching process. From this theory, we know that invasion fitness  $\rho(z_m, z)$  is given by the leading eigenvalue of matrix  $\mathbf{W}(z_m, z)$  (i.e. the mutant goes extinct with probability one if and only if this eigenvalue is less or equal to one [50]). While direct analysis of this eigenvalue is possible (at least numerically), it does not afford much interpretation as is.

To gain greater biological traction, we first introduce the right eigenvector  $\mathbf{q}(z_m, z)$  of  $\mathbf{W}(z_m, z)$ , normalised such that its entries sum to one ( $\sum_{j=1}^M q_j(z_m, z) = 1$ ), in which case  $q_j(z_m, z)$  corresponds to the asymptotic fre-



quency of mutants in class  $j$ . Second, we let  $\mathbf{v}^\circ$  be the left eigenvector of the mean matrix  $\mathbf{W}^\circ = \mathbf{W}(z, z)$  under neutrality (whose  $(i, j)$ -entry gives the expected number of individuals in class  $i$  produced by an individual in class  $j$  in the resident population at demographic equilibrium). Throughout, a superscript  $\circ$  indicates neutrality, i.e. when  $z_m = z$ . Quantities with a superscript  $\circ$  should thus be read as functions of the resident trait  $z$  (e.g.  $\mathbf{v}^\circ$  is a function of  $z$ ) but we do not write such dependency explicitly to avoid notational clutter. We normalise  $\mathbf{v}^\circ$  such that  $\mathbf{v}^\circ \cdot \mathbf{q}^\circ = 1$ , where  $\mathbf{q}^\circ$  is the right eigenvector of the neutral mean matrix  $\mathbf{W}^\circ$ . The  $i$ -entry of the left eigenvector  $\mathbf{v}^\circ$  can then be thought of as the “normalised reproductive value” of an individual in class  $i$ : it is its relative asymptotic demographic contribution to the future of the population in the absence of selection (this left eigenvector ensures that the invasion fitness of a neutral mutant is equal to one, i.e. that  $\rho(z, z) = 1$ , Appendix B.1.1 for more details).

With the above notation, it turns out that the selection gradient can be expressed as

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial w_{ij}(z_m, z)}{\partial z_m} q_j^\circ \quad (10)$$

(here and hereafter when  $s(z)$  is on the left-hand-side of an equation, the derivatives on the right hand side are evaluated at the resident,  $z_m = z$ ; e.g. [23, 27, 72, 73]; Appendix B.1.2 here for derivation). Eq. (10) is most easily read from right to left, starting with  $q_j^\circ$  which is the probability that a randomly sampled individual from a resident lineage (i.e. whose members express  $z$ ) is in class  $j$ . The fitness derivative meanwhile is the effect of a substitution from resident to mutant trait in an individual of class  $j$  on the expected number of offspring in class  $i$  produced by this individual (including itself if it survives and changes class when  $i \neq j$ ). Finally, each offspring is weighted by its reproductive value  $v_i^\circ$ , which is its asymptotic contribution to the future of the population (and thus takes into account the demographic consequences of such offspring).

The implications of eq. (10) are intuitive: selection favours most the expression of traits that increase the production of offspring with high reproductive value in individuals that are more common. Take for instance a situation in which individuals are of either high or low condition. Suppose this is randomly determined at birth with the probability of low condition being greater (so that low condition is more common), but that individuals in high condition have greater fecundity (so that they have greater reproductive value). Gradual evolution in this case will tend to shape traits that favour most the production of high condition offspring by low condition individuals. Conversely, selection is weakest on traits that increase the production of low condition offspring by high condition individuals in this scenario. In the context of social interactions, these considerations and eq. (10) are most relevant in the study of asymmetric games where class determines the payoff consequences of different actions by different players [e.g. 74, 75].

Once a population expresses a singular strategy  $z^*$ , selection may be stabilising or disruptive depending on

$h(z^*)$ , which we show in Appendix B.1.3 can be decomposed as the sum of two biologically relevant terms,

$$h(z^*) = \underbrace{h_w(z^*)}_{\text{non-linear fitness effects}} + 2 \times \underbrace{h_q(z^*)}_{\text{context} \times \text{fitness synergy}}, \quad (11)$$

with

$$\begin{aligned} h_w(z^*) &= \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial^2 w_{ij}(z_m, z)}{\partial z_m^2} q_j^\circ \\ h_q(z^*) &= \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial w_{ij}(z_m, z)}{\partial z_m} \times \frac{\partial q_j(z_m, z)}{\partial z_m}, \end{aligned} \quad (12)$$

where here and hereafter, the derivatives and other terms that characterise the components of  $h(z^*)$  are evaluated at the singular resident  $z_m = z = z^*$  [see 23, 41, 76, for other ways of expressing quadratic selection in matrix populations models]. The first term,  $h_w(z^*)$ , is conceptually equivalent to disruptive selection in an homogeneous population (eq. 8), saying that selection tends to be disruptive when the fitness of an individual increases more than linearly with the expression of its own trait. Under class-structure however, these effects depend on the frequency of the individuals that express them and the reproductive value of their offspring (as in eq. 10). The second term of eq. (11),  $h_q(z^*)$ , is unique to class-structured populations. It reveals that disruptive selection may be driven by the combined effects of a trait change on: (i) the fitness of a focal individual in a given class (say  $j$ ,  $\partial w_{ij}(z_m, z)/(\partial z_m)$ ); and (ii) on the probability that the mutant causing this trait change is in an individual in that class ( $\partial q_j(z_m, z)/(\partial z_m)$ ). More intuitively perhaps,  $h_q(z^*)$  says that disruptive selection may occur when carrying the mutant simultaneously increases (i) the probability of being in a certain class say  $j$ ; and (ii) fitness in that class  $j$ . Disruptive selection may thus be driven by synergistic effects between the context in which the mutant is expressed and fitness in that context. The polymorphism that emerges under such disruptive selection is expected to lead to the coexistence of different morphs that specialise in different classes. This will be made more explicit in the next section where we focus on age-structure.

### 3.2 Age-structure

For many animals, especially endotherms like humans, a major axis of variation is age which is associated with many physiological, behavioural and morphological differences [23, 48, 77]. In discrete time, evolution in age-structured populations can be modelled using the formalism summarised in the preceding section. Since offspring are all born in the same age and age increases linearly with time, the matrix  $\mathbf{W}(z_m, z)$  in eq. (9) becomes a so-called Leslie matrix (Appendix B.2.1), whose special form leads to further insights that we review here.

We first introduce some notation to describe evolution in an age-structured population. Let  $b_a(z_m, z)$  be the expected number of offspring of age 1 produced by a mutant of age  $a \in \{1, \dots, M\}$  with trait  $z_m$  in a resident population with trait  $z$  (where  $M$  is maximum age);  $p_a(z_m, z)$  be the probability that a mutant survives from age  $a$  to

$a + 1$  (so that  $\mu_a(z_m, z) = 1 - p_a(z_m, z)$  is the probability that it dies);  $l_a(z_m, z) = p_1(z_m, z)p_2(z_m, z) \dots p_{a-1}(z_m, z)$  be the probability that a mutant survives at least until age  $a$ ;  $T^\circ = \sum_{a=1}^M a l_a^\circ b_a^\circ$  be the generation time in a population monomorphic for  $z$ , i.e. the expected age of a parent; and finally,

$$\tilde{v}_a^\circ = \sum_{k=a}^M \frac{l_k^\circ}{l_a^\circ} b_k^\circ \quad (13)$$

be the expected number of offspring that an individual produces over the rest of its lifetime given it has survived to age  $a$  in a population monomorphic for  $z$ . This  $\tilde{v}_a^\circ$  is proportional to normalised reproductive value in age-structured populations,  $v_a^\circ$  (specifically,  $v_a^\circ = (L^\circ / T^\circ) \tilde{v}_a^\circ$ , where  $L^\circ = \sum_{j=1}^M l_j^\circ$ , is the expected lifespan of a resident, eq. B-43 in Appendix B.2.2 for details). In fact,  $\tilde{v}_a^\circ$  is often simply referred to as “reproductive value” following Fisher’s seminal work [78]. To distinguish between both definitions, we refer to  $\tilde{v}_a^\circ$  as “current reproductive value” here.

Using the above notation, the selection gradient can be expressed as,

$$s(z) = \frac{1}{T^\circ} \sum_{a=1}^M \left[ \frac{\partial b_a(z_m, z)}{\partial z_m} - \tilde{v}_{a+1}^\circ \frac{\partial \mu_a(z_m, z)}{\partial z_m} \right] l_a^\circ \quad (14)$$

([48, 79]; Appendix B.2.3 here for derivation). This shows that as expected, selection favours an increase in the fecundity ( $b_a(z_m, z)$ ) and a decrease in mortality ( $\mu_a(z_m, z)$ ) at each age  $a$ . The strength of selection on these age-specific fitness components however is proportional to the probability of surviving till that age under neutrality,  $l_a^\circ$ , which can be thought of as the probability that the effect of a mutant at age  $a$  is expressed and thus exposed to selection ( $l_a^\circ$  is in fact proportional to  $q_a^\circ$ , the probability that a resident individual is of age  $a$ , eq. B-36 in Appendix B.2.2 for connection). Since  $l_a^\circ$  always decreases with  $a$ , selection tends to be weaker on later acting than on early acting mutants [79]. This is always true for mutants affecting fecundity ( $b_a(z_m, z)$ ). For mutants affecting mortality ( $\mu_a(z_m, z)$ ), selection strength is further proportional to the remaining number of offspring that an individual is expected to produce if it survives to the next age,  $\tilde{v}_{a+1}^\circ$  (i.e. the current reproductive value, eq. 13). This quantity may in fact increase with age, for instance when maturity occurs later in life. Selection on a mutant that reduces mortality may therefore become stronger as its effects get closer to age at maturity. These well-known results lay the basis of the evolution of life-history traits, especially of senescence [48, 77, 79], and are relevant to age-specific social behaviour [e.g. 80].

Less well-trodden is disruptive selection in an age-structured population, which in fact we have not seen anywhere expressed as eq. (11), together with

$$\begin{aligned} h_w(z^*) &= \frac{1}{T^\circ} \sum_{a=1}^M \left[ \frac{\partial^2 b_a(z_m, z)}{\partial z_m^2} - \tilde{v}_{a+1}^\circ \frac{\partial^2 \mu_a(z_m, z)}{\partial z_m^2} \right] l_a^\circ \\ h_q(z^*) &= \frac{1}{T^\circ} \sum_{a=1}^M \left[ \frac{\partial b_a(z_m, z)}{\partial z_m} - \tilde{v}_{a+1}^\circ \frac{\partial \mu_a(z_m, z)}{\partial z_m} \right] \frac{\partial l_a(z_m, z)}{\partial z_m} \end{aligned} \quad (15)$$

(Appendix B.2.4 for derivation; e.g. [81] for other approaches to disruptive selection in age-structured populations). The term  $h_w(z^*)$  depends on how age-specific fitness components change non-linearly with trait (with

age-specific effects weighted accordingly, as in eq. 14). The second term  $h_q(z^*)$  depends on how fecundity and mortality change with trait expression at a given age  $a$  (the term within square brackets), multiplied to the trait effect on the probability of surviving till that age ( $\partial l_a(z_m, z)/(\partial z_m)$ ). To see the potential relevance of this, consider a scenario where there are two age classes and the evolving trait  $z$  is the effort invested into fecundity at age 1. Suppose that expanding more effort diverts resources from other fitness components, leading to increased mortality at age 1 and decreased fecundity at age 2. An increase in  $z$  thus results in a decrease in both the probability of surviving till age 2 (i.e.  $\partial l_2(z_m, z)/(\partial z_m) < 0$ ) as well as fecundity at that age (i.e.  $\partial b_2(z_m, z)/(\partial z_m) < 0$ ), so that  $\partial l_2(z_m, z)/(\partial z_m) \times \partial b_2(z_m, z)/(\partial z_m) > 0$  causing an increase in  $h_q(z^*)$  and thus to selection being disruptive. In fact, we show in Appendix B.2.5 that polymorphism may emerge in this scenario when individuals compete within age-class. Disruptive selection in that case leads to the coexistence of two highly-differentiated morphs: one that reproduces at age 1 and then dies, and another that reproduces only at age 2. Beyond this specific scenario, eq. (15) reveals how age-structure opens pathways for disruptive selection and thus for the maintenance of genetic variation within populations.

### 3.3 Physiological structure

Age-specific fitness effects of traits are often mediated by some other characteristics of an organism, such as size, knowledge or another (often physiological) attribute, which can depend on an organism's past behaviour, environment, or ontogeny. To characterise trait evolution in such cases, let  $x(a) \in \mathbb{R}$  denote the “internal state” of a mutant individual at age  $a$  and let age  $a \in \mathbb{R}_{\geq 0}$  now be a continuous variable (we use continuous time here as it connects more straightforwardly to existing models and methods). The state  $x(a)$  could be the size of this individual, its foraging skill, or its investment into cooperation at age  $a$ . These individual characteristics develop over time in a way that depends on an individual's traits or behaviours. To model such ontogeny, let us assume all individuals are born with the same initial state  $x(0)$ , which then develops over age according to a differential equation,

$$\frac{dx(a)}{da} = g(z_m, z, x(a)) \quad (16)$$

where the function  $g(z_m, z, x(a))$  gives the rate of change in the internal state of a mutant. This rate of change may depend on the trait  $z_m$  expressed by the mutant, the traits of others it interact with (i.e. the resident  $z$ ), and its current state  $x(a)$ . Eq. (16) also allows the rate of change of the internal state to depend on the entire distribution of resident individuals across states via the resident trait  $z$  (e.g. the distribution of sizes in the resident population). A wide range of models in behavioural ecology and life-history theory can be captured conceptually by eq. (16) (e.g. behavioral response rules, learning rules, neural networks [77, 82–96]). Some of these models conceive the evolving trait  $z$  itself as a function of age or state (e.g. writing  $z(a, x(a))$  and letting this function evolve); we focus here on the case where the trait  $z$  is fixed throughout an individual's lifetime (but see Box 1 for directional selection on age- and state-dependent traits).

The continuous time nature of age calls for several modifications in the way relevant quantities are defined

(in contrast to section 3.2 where age is discrete). First, the fecundity and mortality of an individual at a given age are now rates. These rates can depend on an individual's trait, that of conspecifics, and the individual's current state, and so are written as  $b(z_m, z, x(a))$  and  $\mu(z_m, z, x(a))$  for a mutant at age  $a$  (like eq. 16, these may also depend on the distribution of resident individuals across states through the dependence on resident  $z$ ). Survival  $l(a)$  to age  $a$  is then defined according to a differential equation,

$$\frac{dl(a)}{da} = -\mu(z_m, z, x(a))l(a), \quad \text{with } l(0) = 1. \quad (17)$$

Generation time in the resident monomorphic population now reads as  $T^\circ = \int_0^M a b(z, z, x^\circ(a)) l^\circ(a) da$ , where  $x^\circ(a)$  and  $l^\circ(a)$  are the internal state and survival of resident individuals at age  $a$ , obtained by evaluating eqs. (16) and (17) at monomorphic resident population for  $z$ . We also define

$$\tilde{v}^\circ(a) \equiv \tilde{v}^\circ(a, x^\circ(a)) = \int_a^M \frac{l^\circ(a')}{l^\circ(a)} b(z, z, x^\circ(a')) da' \quad (18)$$

as current reproductive value, i.e. the expected number of offspring that a resident individual produces over the rest of its lifetime given it has survived to age  $a$  and is in state  $x^\circ(a)$ . This is conceptually equivalent to eq. (13), except that here, current reproductive value depends on state  $x^\circ(a)$ . In fact, a change in state  $x^\circ(a)$  at age  $a$  influences current reproductive value at that age,  $\tilde{v}^\circ(a, x^\circ(a))$ , by affecting jointly future survival  $l^\circ(a')$  (via eq. 17, with  $a' > a$ ) and future fecundity  $b(z, z, x^\circ(a'))$  in the (via  $x^\circ(a')$ ). For presentation purposes though, we do not write the dependence of current reproductive value on  $x^\circ(a)$  and use  $\tilde{v}^\circ(a)$  for short (Appendix B.3.2 for details on current reproductive value).

Using these definitions and methods from optimal control theory [92, 95, 97, 98], directional selection on a trait  $z$  that influences how an individual's internal state develops with age (according to eq. 16) can be decomposed as,

$$s(z) = \frac{1}{T^\circ} \int_0^M \left( \frac{\partial b(z_m, z, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(z_m, z, x^\circ(a))}{\partial z_m} + \frac{\partial \tilde{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial g(z_m, z, x^\circ(a))}{\partial z_m} \right) l^\circ(a) da \quad (19)$$

(eq. 19 in [96] for the more general case in group-structured population and Appendix B.3 here for derivation). The first two terms of eq. (19) are conceptually similar to eq. (14), giving directional selection on age-specific fecundity and mortality. More interestingly, the last term within brackets of eq. (19) reveals that selection now also depends on how the trait influences the instantaneous rate of change in internal state ( $\partial g(z_m, z, x^\circ(a))/\partial z_m$ ), and in turn how a change in internal state affects current reproductive value ( $\partial \tilde{v}^\circ(a)/\partial x^\circ(a)$ ). To intuit the relevance of this, it is useful to see  $\partial \tilde{v}^\circ(a)/\partial x^\circ(a)$  as a measure of the fitness value of future reproduction relative to current reproduction owing to a change in internal state at age  $a$ . The last term of eq. (19) then says that in a situation where for example  $x(a)$  is size at age  $a$  and  $z$  controls the investment into growth at each age, selection favours greater investment into growth even at the expense of fecundity when one unit invested into growth yields greater future reproduction relative to that unit invested into current reproduction (i.e. when  $\partial \tilde{v}^\circ(a)/\partial x^\circ(a) \times \partial g(z_m, z, x^\circ(a))/\partial z_m > -\partial b(z_m, z, x^\circ(a))/\partial z_m$ ). In the context of social interactions, eq. (19) would for instance be useful to understand directional selection on re-

active strategies [99, 100], where  $x(a)$  is the level of cooperation at age  $a$  and the trait  $z_m$  determines how an organism reacts to cooperation by its social partners. The last term of eq. (19) in this example would capture how selection at age  $a$  depends on the future benefits of an increase in cooperation at that age.

Disruptive selection, meanwhile, can be decomposed as eq. (11), with

$$\begin{aligned}
 h_w(z^*) &= \frac{1}{T^\circ} \int_0^M \left( \frac{\partial^2 b(z_m, z, x^\circ(a))}{\partial z_m^2} - \bar{v}^\circ(a) \frac{\partial^2 \mu(z_m, z, x^\circ(a))}{\partial z_m^2} + \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial^2 g(z_m, z, x^\circ(a))}{\partial z_m^2} \right) l^\circ(a) da \\
 h_q(z^*) &= \frac{1}{T^\circ} \int_0^M \left[ \underbrace{\left( \frac{\partial b(z_m, z, x^\circ(a))}{\partial z_m} - \bar{v}^\circ(a) \frac{\partial \mu(z_m, z, x^\circ(a))}{\partial z_m} \right) \frac{\partial l(a)}{\partial z_m}}_{\text{change in age}} + \underbrace{h_{q,x}(a) \frac{\partial x(a)}{\partial z_m}}_{\text{change in state}} \right] da
 \end{aligned} \tag{20}$$

(Appendix B.3.5 for derivation). At a broad level,  $h_w(z^*)$  and  $h_q(z^*)$  in eq. (20) have the same interpretation as  $h_w(z^*)$  and  $h_q(z^*)$  in eqs. (12) or (15):  $h_w(z^*)$  depends on the nonlinear effects of the trait on individual fitness components, while  $h_q(z^*)$  depends on how the trait affects both: (i) the proclivity of having a certain age and internal state; and (ii) fitness when in that state. Both  $h_w(z^*)$  and  $h_q(z^*)$  however contain extra terms in comparison to the scenario where only age matters (eq. 15). This is because in addition to age, an individual's state now also depends on  $x(a)$ , which opens new pathways for disruptive selection. The extra term in  $h_w(z^*)$  in eq. (20) (the last term within the brackets) depends on how fitness changes non-linearly with trait expression via a change in state dynamics. This reveals for instance that disruptive selection may be driven by accelerating effects of a trait change on state dynamics at a certain age ( $\partial^2 g(z_m, z, x^\circ(a)) / (\partial z_m)^2 > 0$  e.g. because an extra unit of resources invested in growth at age  $a$  generates a greater than linear increase in growth rate) when such a change improves current reproductive value ( $\partial \bar{v}^\circ(a) / (\partial x^\circ(a)) > 0$ ).

The first part of  $h_q(z^*)$  in eq. (20), labelled “change in age”, is conceptually equivalent to  $h_q(z^*)$  in eq. (15), i.e. capturing the effect of change in representation in a given age class  $a$  (through  $\partial l(a) / (\partial z_m)$ ) but with internal state dynamics left unchanged (so with  $x(a)$  of a resident:  $x^\circ(a)$ ). The effect of changing internal state is contained in the second term of  $h_q(z^*)$  in eq. (20), labelled “change in state”. This consists of the product between how a trait change influences the internal state at age  $a$ ,  $\partial x(a) / (\partial z_m)$ , with  $h_{q,x}(a)$ , which can be thought of as the second-order fitness effect of a change in internal state at age  $a$  (see eq. II.A in Box II for details). As described in Box II, there are several ways that state can influence fitness in a physiologically structured population, such as via interaction effects between trait and state on vital rates. This suggests that disruptive selection can readily emerge owing to a change in state and its knock-on fitness effects. Such disruptive selection would favour polymorphism in the evolving trait  $z$  and as a result, also in internal state across ages ( $x(a)$ ).

We have illustrated eq. (20) (and eq. II.A in Box II) with examples from life-history such as resource allocation problems as those are the most straightforward applications. But since all vital rates (fecundity  $b$ , mortality  $\mu$  and growth  $g$ ) depend on both mutant and resident traits ( $z_m$  and  $z$ ), eqs. (20) and (II.A) can of course be used to understand disruptive selection on traits that influence social interactions. In fact, since the vital rates may depend on the resident internal state and its distribution across resident individuals, eqs. (20) and (II.A) can be applied to social interactions mediated by internal state, such as where larger individuals are more likely to win

contests for resources.

### 3.4 Disruptive selection in class-structured populations

Together, the equations for disruptive selection that we have collected in this section (eqs. 12, 15, and 20) reveal how there are several alternative non-exclusive paths for a trait to become polymorphic in heterogeneous populations, and how these paths depend on the nature of the heterogeneity. In particular, population heterogeneity creates conditions such that individuals may specialise in different contexts (class, age, size, habitat) when trait expression simultaneously increases the proclivity of being in a certain context and fitness in that context (what we have labelled as context  $\times$  direct synergy in eq. 11). This may help explain within-population diversity in traits, such as life-history, sexual development, or habitat choice, that influence the context an individual finds itself in, as well as diversity in social behaviours that are mediated by such heterogeneity.

## 4 Interactions under limited dispersal: the inevitable rise of relatedness

So far, all the scenarios we have explored assume that individuals interact and compete randomly such that a rare mutant only ever interacts with residents. This facilitates analysis because mutant-mutant interactions can be ignored. In reality, carriers of a rare mutation may often interact with other carriers. This is obviously true for within-family interactions, such as parental care or sib competition, but more generally whenever dispersal is limited (i.e. whenever individuals have a non-zero probability of reproducing close to where they were born [101]). As a consequence of limited dispersal, individuals that are physically closer to one another, and thus more likely to interact, are also more likely to share alleles that are identical-by-descent at homologous loci than individuals sampled at random in the population [102]. This inevitably leads to interactions among rare mutants and to what is referred to as kin selection, which occurs whenever a trait expressed by a focal individual affects the fitness of others who are genetically related to the focal at the loci determining the trait [11, 27, 102, 103].

In this last section, we review directional and disruptive selection when dispersal is limited under the light of kin selection. We consider the case where the population is subdivided among a large (effectively infinite) number of groups which can be arbitrarily small. The main assumption is that these groups are equally connected to one another (so that there is no isolation-by-distance): if an individual disperses and leaves its natal group, it is equally likely to immigrate into any other group (as in the homogeneous island model of dispersal of [104], and see [105], for its ecological equivalent). We focus on where there are no exogenous differences among groups (e.g. no differences in environmental condition, [106] for an analysis of this).

The simplest model is where groups are all of the same fixed size  $n$ , and where other than expressing the mutant or resident trait, individuals within groups are homogeneous. We detail the life cycle events that the model can consider in Appendix C.1.1 and simply note here that as long as groups have constant size, variation in several life cycle events are allowed, such as overlapping *vs* non-overlapping generations and dispersal via “migrant

pool” *vs* “propagule pool“ model (in the former, individuals disperse independently from one another whereas in the latter, individuals disperse in groups as part of a propagule [107]). Describing selection in this model requires an individual fitness function that takes into account group structure. We let  $\omega(z_*, \bar{z})$  be this function, which gives the expected number of offspring produced by a focal individual with trait  $z_* \in \{z_m, z\}$  over one demographic time period, when its groups neighbours on average express  $\bar{z}$  (all groups other than the one in which the focal individual lives can be considered monomorphic for the resident  $z$  in an invasion analysis but we do not write this dependency on  $z$  in  $\omega(z_*, \bar{z})$  for simplicity). In writing fitness in terms of the average trait among its neighbours, we are assuming that the focal plays the field within groups. As a more general alternative, fitness may depend on the trait of each individual neighbour, in which case  $\omega(z_*, \bar{z})$  is a first-order approximation in  $\epsilon = z_m - z$  to this more complicated fitness function (Appendix C.1.2 for details).

Invasion fitness can be expressed in terms of the individual fitness function  $\omega(z_*, \bar{z})$  (eq. C-12 in Appendix C.1.2), from which we readily obtain the selection gradient,

$$s(z) = \underbrace{\frac{\partial \omega(z_*, \bar{z})}{\partial z_*}}_{\text{direct effect, } -C} + \underbrace{R^\circ}_{\text{relatedness}} \times \underbrace{\frac{\partial \omega(z_*, \bar{z})}{\partial \bar{z}}}_{\text{indirect effect, } B}, \quad (21)$$

where  $R^\circ$  is the probability that in a population monomorphic for the resident  $z$ , an individual randomly sampled among the neighbours to a random focal individual belong to the same lineage as the focal (i.e. are identical-by-descent, Appendix C.1.3 for derivation).  $R^\circ$  thus corresponds to the standard coefficient of pairwise relatedness [27]. We can recognise in eq. (21) the well-known inclusive fitness effect or Hamilton’s rule in gradient form [11, 20, 27, 29, 30]: the sum of (i) the direct fitness effect, i.e. the effect of a trait change in a focal individual on its own fitness (which in a well-mixed population is the only effect that matters, eq. 8, and which in Hamilton’s rule is typically written as a cost  $-C$ ); and (ii) relatedness-weighted indirect fitness effect, i.e. the effect of a trait change in neighbours on focal fitness (written as a benefit  $B$  in Hamilton’s rule), weighted by the probability that a neighbour and the focal both carry the same mutation (under neutrality). Relatedness in eq. (21) thus quantifies mutant-mutant interactions and highlight their well-known evolutionary significance: interactions among relatives tend to favour the evolution of prosocial traits (i.e. traits such that  $\partial \omega(z_*, \bar{z}) / (\partial \bar{z}) > 0$ , [11, 27, 102]). The selection gradient eq. (21), which is written in terms of the individual fitness function  $\omega(z_*, \bar{z})$  where individuals play the field within groups, also holds more generally for the case where individual fitness depends on the trait of each individual neighbour (Appendix C.1.3 for details).

For group-structured populations, the selection gradient as in eq. (21) is significantly easier to handle mathematically than working from first principles with invasion fitness (or other proxies such as the metapopulation number, [108, 109]). This is because invasion fitness depends on the entire probability distribution that a mutant is in a group with a given number of other mutants ([32, 40, 44, 108, 109]; eq. C-5 in Appendix here). Eq. (21), by contrast, depends only on neutral relatedness, which is just one moment of this distribution assuming the mutant and resident have the same traits. There are at least two advantages to this. First there exist standard techniques coming from coalescent theory to compute such relatedness coefficient ([27]; Ap-



pendix C.1.6 here for an example of such argument). Second, it is perhaps the only evolutionary parameter presented in this review that can be easily and systematically estimated in natural populations. In fact,  $R^\circ$  in eq. (21) can be connected to the well-known  $F_{ST}$  measure of genetic differentiation [110] which can be estimated from neutral markers [111]. Owing to its simplicity, empirical connections, and the biological insights it affords, eq. (21) has been one of the most widely used expression to understand directional selection on an array of social behaviours, such as cooperation, sex-ratio, dispersal and many more [29].

More seldom seen is disruptive selection in terms of relatedness coefficient [44, 46, 106, 112–116]. Under the assumption that individual fitness can be written as  $\omega(z_\bullet, \bar{z})$ , we show in Appendix C.1.4 that disruptive selection is given by,

$$h(z^*) = \underbrace{h_w(z^*)}_{\text{non-linear fitness effects}} + 2 \times \underbrace{h_r(z^*)}_{\text{social context} \times \text{fitness synergy}} \quad (22)$$

where

$$\begin{aligned} h_w(z^*) &= \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet^2} + 2R^\circ \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}} + K^\circ \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial \bar{z}^2} \\ h_r(z^*) &= \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}} \frac{\partial R(z_m, z)}{\partial z_m} \end{aligned} \quad (23)$$

in which  $K^\circ$  is the probability that two individuals, randomly sampled with replacement among the neighbours to a focal individual, are identical-by-descent to the focal (in a population monomorphic for the resident  $z$ ).  $K^\circ$  can thus be thought of as the tendency of interacting with more than one relative under neutrality. The quantity  $R(z_m, z)$ , meanwhile, is the probability that a randomly sampled neighbour to a mutant individual with trait  $z_m$  in a resident population with trait  $z$  is also mutant (note how this measure of genetic structure is no longer under neutrality and depends on the mutant trait  $z_m$ ). If individual fitness more generally depends on the trait of each individual neighbour (so if individuals do not play the field and individual fitness cannot be written simply as  $\omega(z_\bullet, \bar{z})$ ), then the first term  $h_w(z^*)$  consists of extra terms (as in e.g. [44, 113, 114]; Appendix C.1.5 for details).

Eq. (22) highlights how disruptive selection can emerge from two pathways in group-structured populations. The first, given by  $h_w(z^*)$ , depends on three second-order effects of traits on fitness among relatives: (i)  $\partial^2 \omega(z_\bullet, \bar{z}) / (\partial z_\bullet)^2$  is how focal fitness changes non-linearly with its own trait (as in well-mixed populations, eq. 3); (ii)  $\partial^2 \omega(z_\bullet, \bar{z}) / (\partial z_\bullet \partial \bar{z})$ , weighted by relatedness  $R^\circ$ , is how focal fitness changes with joint changes in its own trait and in the average trait among its neighbours; and (iii)  $\partial^2 \omega(z_\bullet, \bar{z}) / (\partial \bar{z})^2$ , weighted by  $K^\circ$ , is how focal fitness changes non-linearly with the average trait in neighbours. These two latter terms capture how trait expression by different individuals within groups can influence focal fitness in a synergistic way (synergy among the focal and the average neighbour with  $\partial^2 \omega(z_\bullet, \bar{z}) / (\partial z_\bullet \partial \bar{z})$ ; and synergy among two average neighbours with  $\partial^2 \omega(z_\bullet, \bar{z}) / (\partial \bar{z})^2$ , [46] for further considerations on these). To see the possible relevance of such synergy, consider for example a scenario where individuals can cooperate with one another and the evolving trait is the amount invested into cooperation and that joint cooperation has antagonistic effects on payoff

(such that  $\partial^2\omega(z_*, \bar{z})/(\partial z_* \partial \bar{z}) < 0$  as in e.g. the snowdrift game). In this case, interactions among relatives tend to inhibit disruptive selection (i.e.  $h(z^*)$  decreases with  $R^\circ$  [114]). Put differently, kin selection favours the evolution of equal contribution among social partners in this scenario (Appendix C.1.6 for analysis; [46, 112, 113] for similar inhibitory effect of limited dispersal in other models).

The second pathway that can contribute to disruptive selection in group-structured populations,  $h_r(z^*)$ , depends on the product of two quantities: (i) how a trait change in neighbours increases the fitness of the focal individual ( $\partial\omega(z_*, \bar{z})/(\partial \bar{z})$ ); with (ii) how a trait change increases the probability of interacting with other individuals also expressing this change, i.e. with relatives ( $\partial R(z_m, z)/(\partial z_m)$ ). This reveals that selection favours mutants that either: (a) increase the fitness of neighbours and the probability that these neighbours are also mutants; or (b) decrease the fitness of neighbours and the probability that these neighbours are also mutants. This effect of selection can thus be seen as the social equivalent of synergy among context and fitness obtained in heterogeneous populations ( $h_q(z^*)$  in eq. 11). In contrast to section 3 where context is the individual state (e.g. sex, age, size or habitat) the mutant can be in, context here is the social environment, i.e. the frequency of relatives in the group. To illustrate such synergy, we explore in Appendix C.1.7 an example where the evolving trait is the amount invested into a common good that benefits the whole group but that such an investment leads to a decreased propensity to disperse (e.g. due to functional trade-offs). As a result, a mutant that invests more resources into cooperation disperses less and thus is more likely to interact with relatives. We show that such a scenario readily leads to the emergence of two social morphs: one that cooperates and tends to remain philopatric, and another that defects and disperses. These two are maintained due to the association between social and dispersal behaviours allowing cooperators to preferentially benefit relatives and defectors to preferentially harm non-relatives. Beyond this specific example,  $h_r(z^*)$  suggests that when mutants can preferentially interact with other mutants (or residents), disruptive selection favours diversity in social behaviours [44, 115, 117–119].

Many social groups are not homogeneous. Colonies of eusocial insects have queens and workers, matriarchal societies of killer whales are composed of multiple generations, and primate groups are often governed by complex dominance hierarchies. We discuss directional and disruptive selection in populations that are subdivided into social groups, and where individuals belong different classes within groups, in boxes III and IV, respectively. The expressions for selection in such heterogeneous social groups combine those of models of class- (eqs. 10-12) and group-structure (eqs. 21-23). In particular, disruptive selection can emerge owing to synergy of fitness with asocial as well as social context (Box IV), laying the ground for the coexistence of morphs that specialise in both types of contexts and thus for adaptive polymorphisms of many different natures.

## 5 Concluding remarks

Understanding phenotypic evolution when individuals interact has been at the core of evolutionary game theory and of the theory of adaptive dynamics. These research programs have led to a well-established and robust

set of tools based on evolutionary invasion analyses that can tackle a wide range of questions in evolutionary biology. Here, we reviewed two important pieces of this toolbox, directional  $s(z)$  and disruptive  $h(z)$  selection, which determine whether gradual evolution leads a population to an uninvadable (i.e. evolutionarily stable) monomorphic state where all individuals express the same trait, or to become polymorphic through adaptive diversification (Fig. 1). As such, they constitute a useful platform to understand the conditions that favour phenotypic variation in the form of adaptive polymorphism.

With the aim of facilitating connections between models and biological interpretation, we collected together expressions of  $s(z)$  and  $h(z)$  in terms of individual fitness, extending reviews that focus on directional selection (e.g. [27, 29–31] for  $s(z)$  in structured populations, including under isolation-by-distance). Interpreting disruptive selection in terms of individual fitness brings together what we see as the strong points of the related branches of theoretical evolutionary biology that are evolutionary game theory and adaptive dynamics. The former has a long tradition of decomposing selection in a way to understand the different forces at play in the evolution of social and life-history traits [5, 6, 27, 29, 79, 102, 120]. Most investigations of complex populations, however, stop at directional selection and do not determine whether polymorphism emerges. By contrast, adaptive dynamics models typically study such emergence as well as its maintenance (how many morphs, the traits they express, and their frequency, e.g. [9, 118, 121–124]). But the complexity of the mathematical procedures involved in such studies and the fitness measures used often leave little room for interpretation. In decomposing disruptive selection in biologically relevant components, we thus hope to motivate yet further research to understand phenotypic variation in non-homogeneous populations, especially in life-history traits and social behaviours that act either via direct interactions, or through indirect ecologically-mediated interactions.

Linking invasion fitness, which is a measure of fitness at the level of the gene (or replicator) [33], to individual fitness in heterogeneous and non-randomly mixed populations requires taking into account the fact that a gene may find itself in carriers in different states, who can in turn interact with other carriers and impact their reproductive success. This is achieved by expressing invasion fitness of a mutant as a weighted average of individual fitness over the distribution of states that a carrier of the mutant can be in ([32, 40, 120]; Appendix C here). With class-structure, the appropriate weights turn out to be reproductive values [72], while the averaging in group-structured populations invariably leads to the notion of relatedness as a measure of interactions among carriers of genes that are identical-by-descent [27, 102]. Remarkably, it is sufficient to consider these summary statistics in the resident population (i.e. under neutrality) to investigate the direction of selection  $s(z)$ : “a gift from God” [125] that bypasses many computational headaches. To characterise disruptive selection  $h(z)$ , however, requires considering how the mutant perturbs the distribution of states an individual carrier can be in (e.g. [44, 106, 113, 114]). As we hope to have conveyed here, these perturbations inform on the nature of disruptive selection. In fact, we suggest that there are two main pathways that can favour polymorphism in trait expression: (i) when individual fitness increases more than linearly with trait expression ( $h_w(z)$ , which is the only pathway in well-mixed and homogeneous populations); (ii) when trait expression simultaneously

increases the probability that an individual is in a given context (e.g. a given age, sex, habitat, size or social environment) and fitness in that context ( $h_q(z)$  and  $h_r(z)$ ; Fig. 2 for summary). Population-structure thus opens novel pathways for the operation of disruptive selection, leading to polymorphisms that have characteristics unique to these populations. More specifically, class- and group-structure lay the ground for the coexistence of genotypes that specialise in the different contexts that a genotype can be in, be it the individual state (e.g. its age, sex or size), or the socio-genetic environment of its carrier (e.g. with or without relatives).

Concerning polymorphism, we have focused our attention on disruptive selection  $h(z)$ , as this constitutes the more recent advances in the literature. It is however important to keep in mind that for phenotypic diversity to emerge, it is also necessary that the population first converges to a singular strategy, i.e. that condition eq. (5) holds. This condition, together with  $h(z^*) > 0$ , highlights that polymorphism requires  $\partial^2 \rho(z_m, z) / (\partial z_m \partial z)$  to be sufficiently negative (when  $z_m = z = z^*$ ). Similar changes in mutant and resident traits must thus lead to a decrease in mutant fitness at the singular strategy. More intuitively perhaps, interactions among mutants and residents must have antagonistic effects on mutant fitness, leading to negative-frequency dependence (which for instance is the case when interactions have payoffs that follow the Snowdrift or Hawk-Dove game).

Owing to time and space constraints, many other results that are relevant to the adaptive dynamics of social behaviour had to be left out. We have for instance ignored the effect of changing or stochastic environments and oscillatory or chaotic population dynamics and finite population sizes (e.g. [22, 126–131]). We have also largely left out the influence of trait evolution on ecological or demographic variables that can feedback on selection and lead for instance to evolutionary suicide (though see Appendix A.3; [14, 132–134]). These feedbacks are particularly relevant under limited dispersal as they lead to inter-temporal mutant-mutant interactions (e.g. when individuals deplete local resources and this disproportionately influences relatives in the future through ecological inheritance [135–137]), whose implications are best understood under the light of kin selection (e.g. [135, 138–144]). With regard to physiological and age-structure, we have not addressed the complications that arise when individuals can be born in different initial internal states (e.g. [145]). Finally, following most of adaptive dynamics literature, we have focused on haploid asexuals. Under random mating and additive genetic effects on traits, directional and disruptive selection are not affected by diploidy and sexual reproduction [27, 146, 147]. This is because when rare, a mutant allele is only ever found in heterozygotic form under random mating (so that the relevant fitness measure is simply the expected number of heterozygotes produced by a heterozygote). But in the presence of inbreeding, whether due to selfing or mating within dispersal-limited groups [108], a rare mutant allele will also appear in homozygotic form. In that case, the machinery described above for class-structure can readily be used to infer evolution (where heterozygotes and homozygotes for the mutant allele constitute two classes, [27, 30]).

These omissions and many others aside, we hope this review provides a basic introduction on how to model and understand Darwinian evolution in non-trivial populations, where density- and frequency-dependence interactions inevitably take place. More broadly, the forms of directional and disruptive selection presented here should help see the connections between fundamental branches of evolutionary biology, from life-history

and evolutionary developmental biology, to social evolution and evolutionary ecology, and thus ultimately facilitate further the integration of these branches.

### Acknowledgements

The authors thank two anonymous reviewers for useful comments, and the Swiss National Science Foundation (PCEFP3181243 to CM) for funding. The authors are grateful to Laurent Lehmann for suggesting to use the mean-field approximation to fitness that allowed us to clarify selection in heterogeneous groups, and more generally for so many enlightening discussions over the years.

### References

- [1] Malthus TR. 1798 *An essay on the principle of population as it affects the future improvement of society*. London, J. Johnson.
- [2] Turchin P. 2001 Does population ecology have general laws? *Oikos* **94**, 17–26.
- [3] Metz JAJH. 2011 pp. 193–231. In *Thoughts on the Geometry of Meso-evolution: Collecting Mathematical Elements for a Postmodern Synthesis*, pp. 193–231. Basel: Springer Basel.
- [4] Vincent TL, Brown JS. 2005 *Evolutionary game theory, natural selection, and Darwinian dynamics*. Cambridge University Press.
- [5] Maynard Smith J, Price GR. 1973 The logic of animal conflict. *Nature* **246**, 15–18.
- [6] Maynard Smith J. 1974 The theory of games and the evolution of animal conflicts. *Journal of theoretical biology* **47**, 209–221.
- [7] Parker G, Maynard Smith J. 1990 Optimality theory in evolutionary biology. *Nature* **348**, 27 – 33.
- [8] Metz JA, Geritz SA, Meszéna G, Jacobs FJ, Van Heerwaarden JS. 1996 Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction. In Van Strien SJ, Verduyn Lunel SM, editors, *Stochastic and Spatial Structures of Dynamical Systems* p. 183—231. Dordrecht, Netherlands: North Holland.
- [9] Geritz S, Kisdi E, Meszena G, a J Metz J. 1998 Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* **12**, 35–57.
- [10] Heino M, Metz AJ, Kaitala V. 1998 The enigma of frequency-dependent selection. *TREE* **13**, 367–370.
- [11] Frank SA. 1998 *Foundations of social evolution*. Princeton, NJ: Princeton University Press.
- [12] McNamara JM, Leimar O. 2020 *Game theory in biology: concepts and frontiers*. Oxford University Press.

- [13] Dieckmann U, Doebeli M, Metz JA, Tautz D. 2004 *Adaptive speciation*. Cambridge University Press.
- [14] Dercole F, Rinaldi S. 2008 *Analysis of evolutionary processes: the adaptive dynamics approach and its applications*. Princeton, NJ: Princeton University Press.
- [15] Pásztor L, Botta-Dukát Z, Magyar G, Czárán T, Meszèna G. 2016 *Theory-based ecology : a Darwinian approach*. Oxford University Press.
- [16] Eshel I. 1991 Game theory and population dynamics in complex genetical systems: the role of sex in short term and in long term evolution. In *Game equilibrium models I* pp. 6–28. Springer.
- [17] Eshel I. 1996 On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution.. *Journal of mathematical biology* **34**, 485–510.
- [18] Hammerstein P. 1996 Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology* **34**, 511–532.
- [19] Weissing FJ. 1996 Genetic versus phenotypic models of selection: can genetics be neglected in a long-term perspective? *Journal of Mathematical Biology* **34**, 533–555.
- [20] Cleve JV. 2015 Social evolution and genetic interactions in the short and long term. *Theoretical Population Biology* **103**, 2–26.
- [21] Tuljapurkar S. 1989 An uncertain life: Demography in random environments. *Theoretical Population Biology* **35**, 227–294.
- [22] Ferriere R, Gatto M. 1995 Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. *Theoretical Population Biology* **48**, 126–171.
- [23] Caswell H. 2001 *Matrix population models: construction, analysis, and interpretation*. Sinauer associates, Inc., Sunderland, MA.
- [24] Tuljapurkar S, Horvitz CC, Pascarella JB. 2003 The many growth rates and elasticities of populations in random environments.. *The American Naturalist* **162**, 489–502.
- [25] Otto SP, Day T. 2007 *A biologist's Guide to mathematical modeling in ecology and evolution*. Princeton University Press.
- [26] Geritz SAH, Éva Kisdi, Meszéna G, Metz JAJ, Hordijk L. 2004 Adaptive Dynamics of Speciation: Ecological Underpinnings. *Interim Report for International Institute for Applied Systems Analysis* **68**, 1–23.
- [27] Rousset F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.
- [28] Kisdi E, Geritz SAH. 2010 Adaptive dynamics: A framework to model evolution in the ecological theatre. *Journal of Mathematical Biology* **61**, 165–169.

- [29] Lehmann L, Rousset F. 2010 How life history and demography promote or inhibit the evolution of helping behaviours. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2599–2617.
- [30] Lehmann L, Rousset F. 2014 The genetical evolution of social behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 17–52.
- [31] Lion S. 2016 Moment equations in spatial evolutionary ecology. *Journal of Theoretical Biology* **405**, 46–57.
- [32] Lehmann L, Mullon C, Akçay E, Van Cleve J. 2016 Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution* **70**, 1689–1702.
- [33] Dawkins R. 1976 *The Selfish Gene*. Oxford University Press.
- [34] Dawkins R. 1978 Replicator Selection and the Extended Phenotype. *Zeitschrift für Tierpsychologie* **47**, 61–76.
- [35] Bulmer M. 1994 *Theoretical evolutionary ecology*. Sinauer associates, Inc., Sunderland, MA.
- [36] Bulmer MG, Taylor. 1980 Local mate competition and the sex ratio. *Theoretical Population Biology* **86**, 406–419.
- [37] Motro U. 1982 Optimal rates of dispersal I. Haploid populations. *Theoretical Population Biology* **21**, 394–411.
- [38] Bulmer MG. 1986 Sex-ratio theory in geographically structured populations. *Heredity* **56**, 69–73.
- [39] Courteau J. 1999 Stochastic Effects in LMC Models. *Theoretical Population Biology* **55**, 127–136.
- [40] Wild G. 2011 Direct fitness for dynamic kin selection. *Journal of Evolutionary Biology* **24**, 1598–610.
- [41] Williams PD, Kamel SJ. 2021 Evolutionary Invasion Analysis in Structured Populations. *Evolutionary Biology* **48**, 422–427.
- [42] Leimar O. 2009 Multidimensional convergence stability. *Evolutionary Ecology Research* **11**, 191–208.
- [43] Débarre F, Nuismer SL, Doebeli M. 2014 Multidimensional (Co)Evolutionary Stability. *The American Naturalist* **184**, 158–171.
- [44] Mullon C, Keller L, Lehmann L. 2016 Evolutionary stability of jointly evolving traits in subdivided populations. *The American Naturalist* **188**.
- [45] Geritz SAH, Metz JAJ, Rueffler C. 2016 Mutual invadability near evolutionarily singular strategies for multivariate traits, with special reference to the strongly convergence stable case. *Journal of Mathematical Biology* **72**, 1081–1099.
- [46] Mullon C, Lehmann L. 2019 An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* **73**, 1695–1728.

- [47] Metz JA, Nisbet RM, Geritz SA. 1992 How should we define ‘fitness’ for general ecological scenarios? *Trends in ecology & evolution* **7**, 198–202.
- [48] Charlesworth B. 1994 *Evolution in age-structured populations* vol. 2. Cambridge University Press Cambridge.
- [49] Rand DA, Wilson HB, McGlade JM. 1994 Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Philosophical Transactions of the Royal Society B-Biological Sciences* **343**, 261–283.
- [50] Harris TE. 1963 *The theory of branching processes* vol. 6. Courier Corporation.
- [51] Kimura M. 1965 A stochastic model concerning the maintenance of genetic variability in quantitative character. *Proceedings of the National Academy of Sciences of the United States of America* **54**, 731–736.
- [52] Fleming WH. 1979 Equilibrium distributions of continuous polygenic traits. *SIAM Journal on Applied Mathematics* **36**, 148–168.
- [53] Lande R. 1979 Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution* **33**, 402.
- [54] Burger R. 1986 On the maintenance of genetic variation: global analysis of Kimura’s continuum-of-alleles model. *Journal of Mathematical Biology* **24**, 341–351.
- [55] Geritz SA. 2005 Resident-invader dynamics and the coexistence of similar strategies. *Journal of Mathematical Biology* **50**, 67–82.
- [56] Meszéna G, Gyllenberg M, Jacobs FJ, Metz JA. 2005 Link between population dynamics and dynamics of Darwinian evolution. *Physical Review Letters* **95**, 078105.
- [57] Rousset F 2006 Separation of time scales, fixation probabilities and convergence to evolutionarily stable states under isolation by distance. *Theoretical Population Biology* **69**, 165–179.
- [58] Dercole F, Geritz SA. 2016 Unfolding the resident-invader dynamics of similar strategies. *Journal of Theoretical Biology* **394**, 231–254.
- [59] Priklopil T, Lehmann L. 2020 Invasion implies substitution in ecological communities with class-structured populations. *Theoretical Population Biology* **134**, 36–52.
- [60] Cai Y, Geritz SA. 2020 Resident-invader dynamics of similar strategies in fluctuating environments. *Journal of Mathematical Biology* **81**, 907–959.
- [61] Eshel I, Motro U. 1981 Kin selection and strong evolutionary stability of mutual help. *Theoretical population biology* **19**, 420–433.
- [62] Eshel I. 1983 Evolutionary and continuous stability. *Journal of Theoretical Biology* **103**, 99–111.



- [63] Taylor PD. 1989 Evolutionary stability in one-parameter models under weak selection. *Theoretical Population Biology* **36**, 125–143.
- [64] Christiansen FB. 1991 On conditions for evolutionary stability for a continuously varying character. *The American Naturalist* **138**, 37–50.
- [65] Eshel I, Motro U, Sansone E. 1997 Continuous stability and evolutionary convergence. *Journal of Theoretical Biology* **185**, 333–343.
- [66] Rueffler C, Dooren TJMV, Leimar O, Abrams PA. 2006 Disruptive selection and then what? *Trends in Ecology and Evolution* **21**, 238–245.
- [67] Parker GA. 1970 The reproductive behaviour and the nature of sexual selection in *Scatophaga stercoraria* L. (*Diptera: Scatophagidae*): II. The fertilization rate and the spatial and temporal relationships of each sex around the site of mating and oviposition. *The Journal of Animal Ecology* **39**, 205.
- [68] Charnov EL. 1976 Optimal foraging, the marginal value theorem. *Theoretical Population Biology* **9**, 129–136.
- [69] Lande R, Arnold S. 1983 The measurement of selection on correlated characters. *Evolution* **37**, 1210–1226.
- [70] Phillips PC, Arnold SJ. 1989 Visualizing multivariate selection. *Evolution* **43**, 1209–1222.
- [71] McGill BJ, Brown JS. 2007 Evolutionary game theory and adaptive dynamics of continuous traits. *Annual Review of Ecology, Evolution, and Systematics* **38**, 403–435.
- [72] Taylor PD. 1990 Allele-frequency change in a class-structured population. *The American Naturalist* **135**, 95–106.
- [73] Barfield M, Holt RD, Gomulkiewicz R. 2011 Evolution in stage-structured populations. *The American Naturalist* **177**, 397–409.
- [74] Frank SA. 1996 Policing and group cohesion when resources vary. *Animal Behaviour* **52**, 1163–1169.
- [75] Rodrigues AM, Gardner A. 2013 Evolution Of Helping And Harming In Heterogeneous Groups. *Evolution* **67**, 2284–2298.
- [76] Shyu E, Caswell H. 2014 Calculating second derivatives of population growth rates for ecology and evolution. *Methods in Ecology and Evolution* **5**, 473–482.
- [77] Stearns S. 1992 *The evolution of life histories*. Oxford: Oxford University Press.
- [78] Fisher R. 1930 *The genetical theory of natural selection*. Clarendon Press.
- [79] Hamilton WD. 1966 The moulding of senescence by natural selection. *Journal of Theoretical Biology* **12**, 12–45.

- [80] Ronce O, Gandon S, Rousset F. 2000 Kin selection and natal dispersal in an age-structured population. *Theoretical Population Biology* **58**, 143–159.
- [81] Rueffler C, Metz JA, van Dooren TJ. 2013 What life cycle graphs can tell about the evolution of life histories. *Journal of Mathematical Biology* **66**, 225–279.
- [82] León JA. 1976 Life histories as adaptive strategies. *Journal of Theoretical Biology* **60**, 301–335.
- [83] Bulmer M. 1981 Worker-queen conflict in annual social Hymenoptera. *Journal of Theoretical Biology* **93**, 239–251.
- [84] J. MJA, O. D. 1986 *The Dynamics of Physiologically Structured Populations*. Springer Berlin Heidelberg.
- [85] Mangel M, Clark CW et al.. 1988 *Dynamic modeling in behavioral ecology*. Princeton University Press.
- [86] Perrin N, Sibly R, Nichols N. 1993 Optimal growth strategies when mortality and production rates are size-dependent. *Evolutionary Ecology* **7**, 576–592.
- [87] Day T, Taylor PD. 1997 Von Bertalanffy's growth equation should not be used to model age and size at maturity. *The American Naturalist* **149**, 381–393.
- [88] Houston AI, McNamara JM et al.. 1999 *Models of adaptive behaviour: an approach based on state*. Cambridge University Press.
- [89] Cichoń M, Kozłowski J. 2000 Ageing and typical survivorship curves result from optimal resource allocation. *Evolutionary Ecology Research* **2**, 857–870.
- [90] Day T, Taylor PD. 2000 A generalization of Pontryagin's maximum principle for dynamic evolutionary games among relatives. *Theoretical Population Biology* **57**, 339–356.
- [91] Irie T, Iwasa Y. 2005 Optimal growth pattern of defensive organs: the diversity of shell growth among mollusks. *The American Naturalist* **165**, 238–249.
- [92] Metz JAJ, Staňková K, Johansson J. 2016 The canonical equation of adaptive dynamics for life histories: from fitness-returns to selection gradients and Pontryagin's maximum principle. *Journal of Mathematical Biology* **72**, 1125–1152.
- [93] Parvinen K, Heino M, Dieckmann U. 2013 Function-valued adaptive dynamics and optimal control theory. *Journal of Mathematical Biology* **67**, 509–533.
- [94] Avila P, Fromhage L, Lehmann L. 2019 Sex-allocation conflict and sexual selection throughout the lifespan of eusocial colonies. *Evolution* **73**, 1116–1132.
- [95] Avila P, Priklopil T, Lehmann L. 2021 Hamilton's rule, gradual evolution, and the optimal (feedback) control of phenotypically plastic traits. *Journal of Theoretical Biology* **526**, 110602.

- [96] Lehmann L. 2022 Hamilton's rule, the evolution of behavior rules and the wizardry of control theory. *Journal of Theoretical Biology* **555**, 111282.
- [97] Liberzon D. 2011 *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press.
- [98] Athans M, Falb PL. 2013 *Optimal control: an introduction to the theory and its applications*. Courier Corporation.
- [99] Killingback T, Doebeli M. 2002 The continuous prisoner's dilemma and the evolution of cooperation through reciprocal altruism with variable investment. *The American Naturalist* **160**, 421–438.
- [100] Lehmann L, Keller L. 2006 The evolution of cooperation and altruism - A general framework and a classification of models. *Journal of Evolutionary Biology* **19**, 1365–1376.
- [101] Clobert J, Baguette M, Benton TG. 2012 *Dispersal ecology and evolution*. Oxford University Press.
- [102] Hamilton WD. 1964 The genetical evolution of social behaviour, I and II. *Journal of Theoretical Biology* **7**, 1–52.
- [103] Michod RE. 1982 The theory of kin selection. *Annual Review of Ecology and Systematics* **13**, 23–55.
- [104] Wright S. 1931 Evolution in Mendelian populations. *Genetics* **16**, 97.
- [105] Chesson PL. 1981 Models for spatially distributed populations: The effect of within-patch variability. *Theoretical Population Biology* **19**, 288–325.
- [106] Ohtsuki H, Rueffler C, Wakano JY, Parvinen K, Lehmann L. 2020 The components of directional and disruptive selection in heterogeneous group-structured populations. *Journal of Theoretical Biology* **507**.
- [107] Slatkin M. 1977 Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical Population Biology* **12**, 253–262.
- [108] Parvinen K, Metz J. 2008 A novel fitness proxy in structured locally finite metapopulations with diploid genetics, with an application to dispersal evolution. *Theoretical Population Biology* **73**, 517–528.
- [109] Massol F, Calcagno V, Massol J. 2009 The metapopulation fitness criterion: Proof and perspectives. *Theoretical Population Biology* **75**, 183–200.
- [110] Rousset F. 2002 Inbreeding and relatedness coefficients: what do they measure? *Heredity* **88**, 371–380.
- [111] Charlesworth B, Charlesworth D. 2010 *Elements of Evolutionary Genetics*. Roberts and Company Publishers.
- [112] Day T. 2001 Population structure inhibits evolutionary diversification under competition for resources. *Genetica* **112-113**, 71–86.

- [113] Ajar. 2003 Analysis of disruptive selection in subdivided populations.. *BMC Evolutionary Biology* **3**, 22.
- [114] Wakano JY, Lehmann L. 2014 Evolutionary branching in deme-structured populations. *Journal of Theoretical Biology* **351**, 83–95.
- [115] Mullon C, Keller L, Lehmann L. 2018 Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nature Ecology and Evolution* **2**, 132–140.
- [116] Mullon C, Lehmann L. 2022 Evolution of warfare by resource raiding favours polymorphism in belligerence and bravery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **377**.
- [117] Koella JC. 2000 The spatial spread of altruism versus the evolutionary response of egoists. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 1979–1985.
- [118] Parvinen K. 2013 Joint evolution of altruistic cooperation and dispersal in a metapopulation of small local populations. *Theoretical Population Biology* **85**, 12–19.
- [119] Joshi J, Åke Brännström, Dieckmann U. 2020 Emergence of social inequality in the spatial harvesting of renewable public goods. *PLOS Computational Biology* **16**, e1007483.
- [120] Taylor PD, Frank SA. 1996 How to make a kin selection model. *Journal of Theoretical Biology* **180**, 27–37.
- [121] Éva Kisdi. 1999 Evolutionary branching under asymmetric competition. *Journal of Theoretical Biology* **197**, 149–162.
- [122] Geritz SA, Meijden EVD, Metz JA. 1999 Evolutionary dynamics of seed size and seedling competitive ability. *Theoretical Population Biology* **55**, 324–343.
- [123] Mathias A, Éva Kisdi. 2002 Adaptive diversification of germination strategies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 151–155.
- [124] Geritz SA, Éva Kisdi, Yan P. 2007 Evolutionary branching and long-term coexistence of cycling predators: critical function analysis. *Theoretical Population Biology* **71**, 424–435.
- [125] Hamilton WD. 2005 *Narrow Roads of Gene Land: Volume 1: Evolution of Social Behaviour*. Spektrum Academic Publishers.
- [126] Proulx S, Day T. 2002 What can invasion analyses tell us about evolution under stochasticity in finite populations? *Selection* **2**, 2–15.
- [127] Cadet C, Ferrière R, Metz JA, Van Baalen M. 2003 The evolution of dispersal under demographic stochasticity. *The American Naturalist* **162**, 427–441.
- [128] Wakano JY, Iwasa Y. 2013 Evolutionary branching in a finite population: deterministic branching vs. stochastic branching. *Genetics* **193**, 229–241.

- [129] Ripa J, Dieckmann U. 2013 Mutant invasions and adaptive dynamics in variable environments. *Evolution* **67**, 1279–1290.
- [130] Débarre F, Otto SP. 2016 Evolutionary dynamics of a quantitative trait in a finite asexual population.. *arXiv* **108**, 75–88.
- [131] Lion S, Gandon S. 2022 Evolution of class-structured populations in periodic environments. *Evolution*.
- [132] Gyllenberg M, Parvinen K. 2001 Necessary and sufficient conditions for evolutionary suicide. *Bulletin of mathematical biology* **63**, 981–993.
- [133] Boldin B, Kisdi E. 2016 Evolutionary suicide through a non-catastrophic bifurcation: adaptive dynamics of pathogens with frequency-dependent transmission. *Journal of mathematical biology* **72**, 1101–1124.
- [134] Lion S. 2018 Theoretical Approaches in Evolutionary Ecology: Environmental Feedback as a Unifying Perspective. *Am. Nat* **191**, 21–44.
- [135] Dawkins R. 1982 *The Extended Phenotype*. Oxford: Oxford University Press.
- [136] Odling-Smee FJ, Laland KN, Feldman MW, Feldman MW. 2003 *Niche Construction: The Neglected Process in Evolution*. Number 37. Princeton University Press.
- [137] Bonduriansky R, Day T. 2020 *Extended heredity: a new understanding of inheritance and evolution*. Princeton University Press.
- [138] Dawkins R. 2004 Extended phenotype—but not too extended. A reply to Laland, Turner and Jablonka. *Biology and Philosophy* **19**, 377–396.
- [139] Lehmann L. 2007 The evolution of trans-generational altruism: kin selection meets niche construction. *Journal of Evolutionary Biology* **20**, 181–189.
- [140] Lehmann L. 2008 The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. *Evolution* **62**, 549–566.
- [141] Lehmann L. 2010 Space-time relatedness and Hamilton's rule for long-lasting behaviors in viscous populations. *American Naturalist* **175**, 136–143.
- [142] Mullan C, Lehmann L. 2018 Eco-evolutionary dynamics in metacommunities: ecological inheritance, helping within species, and harming between species. *The American Naturalist* **192**, 664–686.
- [143] Mullan C, Wakano JY, Ohtsuki H. 2021 Coevolutionary dynamics of genetic traits and their long-term extended effects under non-random interactions. *Journal of Theoretical Biology* **525**.
- [144] Prigent I, Mullan C. 2022 The moulding of intra-specific diversity by selection under ecological inheritance. *bioRxiv*.

- [145] Dieckmann O, Metz JA. 1994 On the reciprocal relationship between life histories and population dynamics. In *Frontiers in Mathematical Biology* pp. 263–279. Springer.
- [146] Geritz SAH, Kisdi E. 2000 Adaptive dynamics in diploid, sexual populations and the evolution of reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences* **267**, 1671–1678.
- [147] Metz J, Leimar O. 2011 A simple fitness proxy for structured populations with continuous traits, with case studies on the evolution of haplo-diploids and genetic dimorphisms. *Journal of Biological Dynamics* **5**, 163–190.
- [148] West-Eberhard MJ. 2003 *Developmental plasticity and evolution*. Oxford University Press.
- [149] Gomulkiewicz R, Kirkpatrick M. 1992 Quantitative genetics and the evolution of reaction norms. *Evolution* **46**, 390–411.
- [150] Parvinen K, Dieckmann U, Heino M. 2006 Function-valued adaptive dynamics and the calculus of variations. *Journal of Mathematical Biology* **52**, 1–26.
- [151] Dieckmann U, Heino M, Parvinen K. 2006 The adaptive dynamics of function-valued traits. *Journal of Theoretical Biology* **241**, 370–389.
- [152] Beder JH, Gomulkiewicz R. 1998 Computing the selection gradient and evolutionary response of an infinite-dimensional trait. *Journal of Mathematical Biology* **36**, 299–319.
- [153] Gomulkiewicz R, Kingsolver JG, Carter PA, Heckman N. 2018 Variation and evolution of function-valued traits. *Annual Review of Ecology, Evolution, and Systematics* **49**, 139–164.
- [154] Frank SA. 2010 A general model of the public goods dilemma. *Journal of Evolutionary Biology* **23**, 1245–1250.
- [155] Rodrigues AM, Gardner A. 2022 Reproductive value and the evolution of altruism. *Trends in Ecology and Evolution* **37**, 346–358.

### Box I. Directional selection on age- and state-dependent expression: the moulding of plastic traits

The selection gradient in eq. (19) gives the fitness effect of a mutant trait  $z_m$  (e.g. the proportion of resources allocated to growth) that influences a fitness-relevant state  $x(a)$  (e.g. size at age  $a$ ) that changes with the age  $a$  of an individual (according to a dynamical system eq. 16). The trait  $z_m$ , however, is assumed to be fixed over an organism's lifetime in eq. (19). A more complicated problem is when the evolving trait  $z$  is itself a function, either of age (so-called “open-loop controls”; e.g. age-dependent resource allocation to growth) or of both age and state (“closed-loop controls”; e.g. size-dependent aggression level). Such traits are more colloquially said to be plastic ([148]; or function-valued, e.g. [118, 149–153]). As it turns out, directional selection on such traits takes a similar form to eq. (19), revealing that singular strategies must satisfy the following balance condition at each age  $a$ :

$$\frac{\partial b(z_m(a), z(a), x^\circ(a))}{\partial z_m(a)} = \bar{v}^\circ(a) \frac{\partial \mu(z_m(a), z(a), x^\circ(a))}{\partial z_m(a)} - \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial g(z_m(a), z(a), x^\circ(a))}{\partial z_m(a)}, \quad (\text{I.A})$$

where  $z_m(a)$  is trait expression at age  $a$  of a mutant (and  $z(a)$  of a resident), which may be written as a function of age only, say  $z_m(a) = u(a)$  for open-loop controls, or of both age and state, say  $z_m(a) = u(a, x(a))$  for closed-loop controls (eq. B-150 in Appendix B.3.7; eq. 29 in [95]). Condition eq. (I.A) reveals for instance that selection favours an increase in current reproduction (left hand side) only if it exceeds the expected loss in future reproduction (right hand side) due to an increase in current mortality and a decrease in the current rate of change of internal state (e.g. increasing fecundity by investing less into cellular repair and growth). In contrast to eq. (19), selection on age  $a$  is independent from the probability of surviving to that age, meaning that where possible, selection favours traits that optimise life-history for each age  $a$ .

Interestingly, strategies that are age- and state-dependent (i.e. open- and closed-loop controls) evolve to produce the same plastic phenotypes in well-mixed populations (given the environment is deterministic) [95]. By contrast, in group-structured populations where individuals interact locally through their state (e.g. size dependent competition for light in plants), age- and state-dependent strategies can lead to different traits. This is because when individuals are able to respond to their own state and that of others, selection favours anticipating the future actions of others (e.g. when growing larger others might respond by growing even larger). This “anticipation” is taken into account in the term  $\partial \bar{v}^\circ(a) / (\partial x^\circ(a))$  in the selection gradient (Appendix B.3.7; [95] for more details; also [40, 90, 95, 96] for models with state-modulated local interactions). We only present results regarding directional selection of plastic traits here as disruptive selection for these traits has not yet been worked out, needing careful consideration of the nature of phenotypic deviation ([151] for further discussion).

## Box II. Second-order fitness effects of a state change in physiologically-structured populations

Disruptive selection in physiologically-structured populations (eq. 20, section 3) depends on the product between the effect of a change in trait expression on internal state at age  $a$  ( $\partial x(a)/(\partial z_m)$ ) and

$$\begin{aligned}
 h_{q,x}(a) = & \left( \frac{\partial b(z, z, x(a))}{\partial x(a)} - \bar{v}^\circ(a) \frac{\partial \mu(z, z, x(a))}{\partial x(a)} \right) \frac{\partial l(a)}{\partial z_m} \\
 & + \left( \frac{\partial^2 b(z_m, z, x(a))}{\partial z_m \partial x(a)} - \bar{v}^\circ(a) \frac{\partial^2 \mu(z_m, z, x(a))}{\partial z_m \partial x(a)} + \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial^2 g(z_m, z, x(a))}{\partial z_m \partial x(a)} \right) l^\circ(a) \\
 & + \frac{1}{2} \left( \frac{\partial^2 b(z, z, x(a))}{\partial x(a)^2} - \bar{v}^\circ(a) \frac{\partial^2 \mu(z, z, x(a))}{\partial x(a)^2} + \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial^2 g(z, z, x(a))}{\partial x(a)^2} \right) \frac{\partial x(a)}{\partial z_m} l^\circ(a),
 \end{aligned} \tag{II.A}$$

where each line corresponds to a different fitness effect of a change in internal state  $x(a)$  at age  $a$ . (i) The first line of  $h_{q,x}(a)$  depends on how a change in state at age  $a$  affects fecundity and mortality at that age, multiplied to the effect of a trait change on the probability of surviving till then ( $\partial l(a)/(\partial z_m)$ ). Consider for instance a model where  $x(a)$  is size and  $z_m$  controls the investment into growth. Under the assumption that growth trades off with survival (so that  $\partial l(a)/(\partial z_m) \times \partial x(a)/(\partial z_m) < 0$ ), this first line multiplied to  $\partial x(a)/(\partial z_m)$  would be positive and thus favour disruptive selection when vital rates decrease with size at the singular strategy (so that the term within brackets in the first line of eq. II.A is negative). (ii) The second line of  $h_{q,x}(a)$  depends on the interaction effects of the evolving trait and internal state on fitness (captured by the cross derivatives with respect to  $z_m$  and  $x(a)$ ). To illustrate the potential implications of this, let us continue with the previous example where  $x(a)$  is size and  $z_m$  controls the investment into growth. Disruptive selection could occur because the fitness cost of investing resources into growth decreases with size (so that the term within brackets in the second line of eq. II.A is negative and its product with  $\partial x(a)/(\partial z_m) < 0$  is positive). (iii) The third line of  $h_{q,x}(a)$  depends on the non-linear effects of a change in state on fitness components. With  $x(a)$  as size for example, the third line would be positive where mortality decreases with body size in an accelerating manner (such that  $\partial^2 \mu(z_m, z, x(a))/(\partial x(a))^2 < 0$ , e.g. because individuals are increasingly better at fending off predators with size). More broadly, eq. (II.A) shows there are multiple ways for state to influence fitness and thus potentially favour disruptive selection.



Consider a population subdivided into social groups and where individuals belong to  $M$  different classes within groups (e.g. age, sex, social rank as in section 3.1). To describe the fitness of a focal individual in this model, we first denote by  $\bar{z}_k$  the average trait expressed among the neighbours to this focal that belong to class  $k$ . For short, we collect these averages in the vector  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_M)$ . We then let  $\omega_{ij}(z_\bullet, \bar{z})$  be the expected number of offspring in class  $i$  produced by a focal individual in class  $j$  with trait  $z_\bullet$  when its group-neighbours express  $\bar{z}$  on average (Appendix C.2.1 for details). As shown previously [27, 32, 120], the selection gradient for this model can be expressed as,

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \left[ \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet} + \sum_{i'=1}^M \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'}} R_{i'|j}^\circ \right] q_j^\circ \quad (\text{III.A})$$

where  $v_i^\circ$  is the reproductive value of individuals in class  $i$ ,  $R_{i'|j}^\circ$  is the probability that a randomly sampled neighbour in class  $i'$  to a focal individual in class  $j$  is identical-by-descent to the focal, and  $q_j^\circ$  is the probability that a randomly sampled individual is in class  $j$  (all three quantities in a population monomorphic for the resident  $z$ ; Appendix C.2.2 for our derivation of eq. III.A where we clarify some of the arguments used by [120]). In a well-mixed population (where  $R_{i'|j}^\circ = 0$  for all  $i'$  and  $j$ ), the selection gradient reduces to the one for class-structured populations (eq. 10), as expected. With limited dispersal and interactions among relatives, selection further depends on indirect fitness effects ( $\partial \omega_{ij}(z_\bullet, \bar{z}) / (\partial \bar{z}_{i'})$ ). As highlighted by eq. (III.A), these indirect fitness effects tend to favour prosocial behaviours, in particular towards individuals that produce offspring with high reproductive value. More generally, eq. (III.A) allows us to understand social evolution within heterogeneous groups and thus under asymmetric interactions which can lead to counter-intuitive situations (i.e. where payoff depends on class, e.g., [74, 75, 154, 155]).

Box IV. Disruptive selection in heterogeneous social groups: where individual and social context can drive polymorphism

Like directional selection (eq. III.A), disruptive selection in heterogeneous social groups combines elements from class- and group-structure (i.e. from eqs. 12 and 23). In fact, we show in Appendix C.2.3 that under the assumption that the expected number of offspring in class  $i$  produced by a focal individual in class  $j$  with trait  $z_{\bullet}$  can be written as  $\omega_{ij}(z_{\bullet}, \bar{z})$ , where  $\bar{z} = (\bar{z}_1, \dots, \bar{z}_M)$  collects the average trait  $\bar{z}_k$  among neighbours in class  $k$  (i.e. under the assumption that individuals play the field within groups), disruptive selection can be decomposed into,

$$h(z^*) = h_w(z^*) + 2h_r(z^*) + 2h_q(z^*). \quad (\text{IV.A})$$

Briefly, the first term consists of second-order fitness effects weighted by reproductive value and relatedness,

$$h_w(z^*) = \sum_{i=1}^M \sum_{j=1}^M v_i^{\circ} \left[ \frac{\partial^2 \omega_{ij}(z_{\bullet}, \bar{z})}{\partial z_{\bullet}^2} + 2 \sum_{i'=1}^M \frac{\partial^2 \omega_{ij}(z_{\bullet}, \bar{z})}{\partial z_{\bullet} \partial \bar{z}_{i'}} R_{i'|j}^{\circ} + \sum_{i'=1}^M \sum_{i''=1}^M \frac{\partial^2 \omega_{ij}(z_{\bullet}, \bar{z})}{\partial \bar{z}_{i'} \partial \bar{z}_{i''}} K_{i',i''|j}^{\circ} \right] q_j^{\circ}, \quad (\text{IV.B})$$

where  $K_{i',i''|j}^{\circ}$  is the probability that in a population monomorphic for the resident  $z$ , two individuals in classes  $i'$  and  $i''$  randomly sampled with replacement among the neighbours to a random focal individual in class  $j$  are identical-by-descent to the focal. The second term,

$$h_r(z^*) = \sum_{i=1}^M \sum_{j=1}^M \sum_{i'=1}^M v_i^{\circ} \frac{\partial \omega_{ij}(z_{\bullet}, \bar{z})}{\partial \bar{z}_{i'}} \frac{\partial R_{i'|j}(z_m, z)}{\partial z_m} q_j^{\circ}, \quad (\text{IV.C})$$

depends on the trait's effect on relatedness, as  $R_{i'|j}(z_m, z)$  is the asymptotic probability that a randomly sampled neighbour in class  $i'$  to a mutant individual in class  $j$  with trait  $z_m$  in a resident population with trait  $z$  is also mutant. Finally, the third term participating to disruptive selection,

$$h_q(z^*) = \sum_{i=1}^M \sum_{j=1}^M v_i^{\circ} \left[ \frac{\partial \omega_{ij}(z_{\bullet}, \bar{z})}{\partial z_{\bullet}} + \sum_{i'=1}^M \frac{\partial \omega_{ij}(z_{\bullet}, \bar{z})}{\partial \bar{z}_{i'}} R_{i'|j}^{\circ} \right] \frac{\partial q_j(z_m, z)}{\partial z_m}, \quad (\text{IV.D})$$

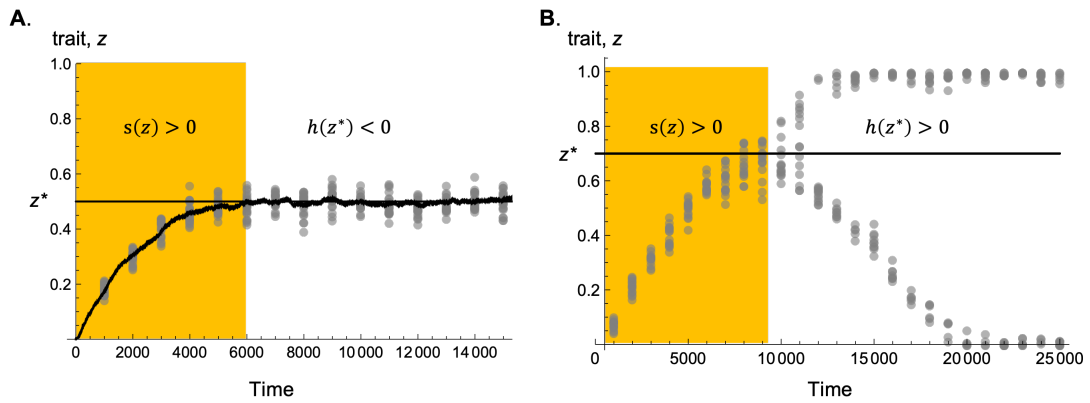
depends on the trait's effect on the probability on being in a certain class,  $\partial q_j(z_m, z) / (\partial z_m)$ . All three terms thus have similar interpretations than those emerging in models of just class- or just group-structure (eqs. 12 and 23). In particular,  $h_w(z^*)$  reveals that disruptive selection can come about when fitness changes non-linearly with trait expression within and between individuals of the same group (as in eq. (23) but here weighted by reproductive value);  $h_r(z^*)$ , when trait expression increases both the likelihood of being in a certain social environment and fitness in that environment; and finally  $h_q(z^*)$ , when trait expression augments jointly the probability of being in a given individual state and fitness in that state (both via direct and indirect fitness effects owing to group-structure).

**Table 1: Key general symbols**

$z_m, z$	Mutant and resident traits, respectively ( $z_m, z \in Z \subseteq \mathbb{R}$ , where $Z$ is the space of all possible strategies).
$\rho(z_m, z)$	Invasion fitness or geometric growth rate of a mutant allele coding for trait $z_m$ in a resident population that is monomorphic for $z$ (i.e. per-capita per-time-step number of mutant copies produced by the mutant lineage, $\rho : Z \times Z \rightarrow \mathbb{R}$ and twice differentiable).
$w(z_m, z)$	Individual fitness of a mutant carrier with trait value $z_m$ when the rest of the population expresses $z$ (i.e. expected number of direct descendants produced over one time step by this individual, $w : Z \times Z \rightarrow \mathbb{R}$ and twice differentiable). In a homogeneous and well-mixed population, $w(z_m, z) = \rho(z_m, z)$ , otherwise not necessarily.
$s(z)$	Directional selection gradient (eq. 3).
$h(z)$	Disruptive selection (eq. 3).
$z^*$	Singular strategy: trait value such that when expressed by the resident population, there is no directional selection (i.e. such that $s(z^*) = 0$ ).

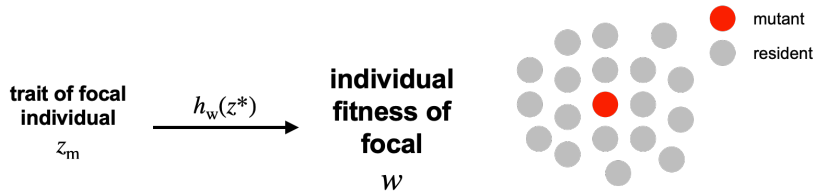
**Table 2: Key symbols for the different models of population structure**

Class-structure (section 3.1)	
$M \in \mathbb{Z}_+$	Number of classes (e.g. $M = 2$ for a model with males and females).
$w_{ij}(z_m, z)$	Expected number of mutants in class $i \in \{1, \dots, M\}$ produced by a mutant in class $j \in \{1, \dots, M\}$ over one time step.
$\mathbf{W}(z_m, z)$	Mean matrix: $M \times M$ matrix with $(i, j)$ -entry $w_{ij}(z_m, z)$ . Invasion fitness $\rho(z_m, z)$ is given by the leading eigenvalue of this matrix.
$\mathbf{q}(z_m, z)$	Asymptotic frequency distribution of mutants across classes (right eigenvector of $\mathbf{W}(z_m, z)$ , normalised such that entries sum to one, $\sum_{j=1}^M q_j(z_m, z) = 1$ ). Denoted by $\mathbf{q}^\circ = \mathbf{q}(z, z)$ under neutrality.
$\mathbf{v}^\circ$	Normalised reproductive values, i.e. $v_i^\circ$ is the relative asymptotic demographic contribution of an individual in class $i$ to the future of the population in the absence of selection. Given by the left eigenvector of $\mathbf{W}^\circ = \mathbf{W}(z, z)$ and normalised such that $\mathbf{v}^\circ \cdot \mathbf{q}^\circ = 1$ .
Age-structure (in discrete time, section 3.2)	
$M \in \mathbb{Z}_+$	Maximum lifespan.
$b_a(z_m, z)$	Fecundity of a mutant at age $a \in \{1, \dots, M\}$ , i.e. expected number of offspring of age 1 produced by a mutant of age $a$ . Under neutrality, $b_a^\circ = b_a(z, z)$ .
$\mu_a(z_m, z)$	Probability of death of a mutant at age $a$ . Under neutrality, $\mu_a^\circ = \mu_a(z, z)$ .
$l_a(z_m, z)$	Probability that a mutant survives at least to age $a$ . Under neutrality, $l_a^\circ = l_a(z, z)$ .
$T^\circ$	Generation time in a population monomorphic for $z$ , i.e. the expected age of a parent.
$\tilde{v}_a^\circ$	Current reproductive value, i.e. expected number of offspring that an individual produces over the rest of its lifetime given it has survived to age $a$ in a population monomorphic for $z$ (eq. 13), proportional to normalised reproductive value, $v_a^\circ = (L^\circ / T^\circ) \tilde{v}_a^\circ$ where $L^\circ = \sum_{j=1}^M l_j^\circ$ , is the expected lifespan of a resident.
Physiological-structure (in continuous-time, section 3.3)	
$M \in \mathbb{R}_+$	Maximum lifespan ( $M = \infty$ when lifespan is endogenously determined).
$x(a), x^\circ(a)$	“Internal states”, or “states” for short (e.g. size, skill), of a mutant and of a resident at age $0 \leq a < M$ , respectively.
$g(z_m, z, x(a))$	Rate of change of the state of a mutant in state $x(a)$ (eq. 16).
$b(z_m, z, x(a))$	Fecundity rate of a mutant in state $x(a)$ .
$\mu(z_m, z, x(a))$	Death rate of a mutant in state $x(a)$ .
$l(a), l^\circ(a)$	Probabilities that a mutant and a resident survive at least until age $a$ , respectively (eq. 17).
$\tilde{v}^\circ(a)$	Current reproductive value, i.e. expected number of offspring that an individual produces over the rest of its lifetime given it has survived to age $a$ in a population monomorphic for $z$ (eq. 18).
Group-structure (section 4)	
$z_\bullet$	Trait of a focal individual ( $z_\bullet \in \{z_m, z\}$ ).
$\bar{z}$	Average trait expressed by the neighbours of the focal individual (i.e. all members of the group except the focal individual).
$\omega(z_\bullet, \bar{z})$	Expected number of offspring produced by the focal individual over one time step.
$R^\circ$	Neutral relatedness: probability that in a population monomorphic for the resident $z$ , an individual randomly sampled among the neighbours to a focal individual belong to the same lineage as the focal (i.e. are identical-by-descent).
$R(z_m, z)$	Mutant relatedness: probability that a randomly sampled neighbour to a mutant individual with trait $z_m$ in a resident population with trait $z$ is also mutant (i.e. are identical-by-descent). Under neutrality, $R(z, z) = R^\circ$ .
$K^\circ$	Probability that two individuals, randomly sampled with replacement among the neighbours to a focal individual, are identical-by-descent to the focal (in a population monomorphic for $z$ ).



**Figure 1: Trait evolution under stabilising and disruptive selection.** **A:** Evolution under recurrent mutations when selection is stabilising. First, the population evolves under directional selection (shaded region) gradually converging to the singular strategy  $z^*$ . Once the population expresses  $z^*$ , stabilising selection ( $h(z^*) < 0$ ) maintains the population monomorphic for  $z^*$ . Simulations are shown for the biological scenario given in Appendix A.1 (parameters used:  $\gamma = 0.0005$ ,  $\mu = 0.8$ ,  $f_0 = 2$ ,  $B_1 = 2$ ,  $B_2 = -2$ ,  $B_3 = 0$ , and with mutations occurring with probability 0.01 in offspring and whose effects on the trait have mean 0 and standard deviation 0.02). Each gray dot is the trait expressed by an individual (we randomly sampled 25 individuals every 1000 generations), thick black line is the population average, and thin black line is the convergence stable and uninvadable strategy  $z^*$ . **B:** Evolution under disruptive selection and the emergence of polymorphism (same model as in **A** with  $B_1 = 1.35$ ,  $B_2 = 0.5$  and  $B_3 = 1$ ). The population first converges to  $z^*$  under directional selection (shaded region) and then becomes dimorphic owing to disruptive selection.

### A. Homogeneous and well-mixed population



### B. Class- and group-structured population

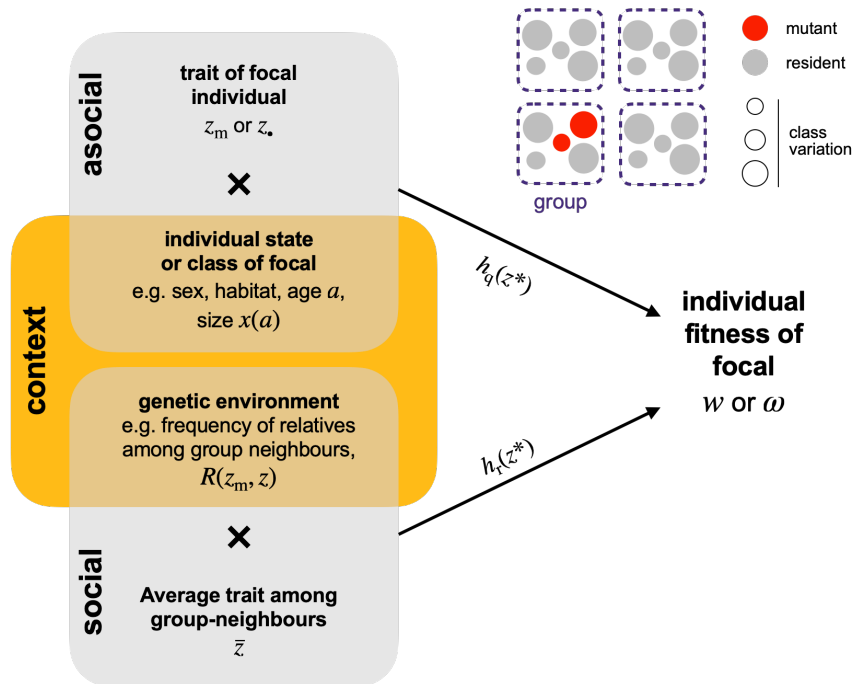


Figure 2: **The different paths for disruptive selection.** **A:** In a homogeneous and well-mixed population, the only relevant effect for disruptive selection is how trait expression by an individual influences its own fitness, specifically whether fitness increases more than linearly with trait expression ( $h_w(z^*)$ , eq. 8). **B:** In class- and/or group-structured populations, disruptive selection can also be due to trait expression simultaneously increasing the probability that a focal individual is in a certain context and individual fitness in that context. Such context  $\times$  fitness synergy can be decomposed in two types: (i) asocial, where the context is the state (or class) of the focal ( $h_q(z^*)$ , eq. 12); and (ii) social, where the context is the genetic environment of the focal ( $h_r(z^*)$ , eq. 23).

# Appendix for “Evolutionary Game theory and the Adaptive Dynamics Approach: Adaptation where Individuals Interact”

Piret Avila<sup>1</sup> and Charles Mullon<sup>2,\*</sup>

<sup>1</sup>Institute for Advanced Studies in Toulouse, Université Toulouse 1 Capitole, 31080 Toulouse, France

<sup>2</sup>Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland

\*Correspondence: [charles.mullon@unil.ch](mailto:charles.mullon@unil.ch)

## Contents

<b>A Well-mixed and homogeneous populations</b>	<b>3</b>
A.1 An example . . . . .	3
A.2 Selection in terms of payoff . . . . .	5
A.3 Connection between selection and demography . . . . .	5
<b>B Selection in class-structured populations</b>	<b>6</b>
B.1 Matrix population models . . . . .	6
B.1.1 Invasion fitness . . . . .	6
B.1.2 Directional selection . . . . .	8
B.1.3 Disruptive selection . . . . .	9
B.2 Age-structure . . . . .	9
B.2.1 Leslie matrix . . . . .	9
B.2.2 Stable age distribution and Fisher’s reproductive value . . . . .	10
B.2.3 Directional selection in age-structured populations . . . . .	11
B.2.4 Disruptive selection in age-structured populations . . . . .	12
B.2.5 Example . . . . .	14
B.3 Physiological structure . . . . .	15
B.3.1 Selection under continuous age and physiological structure . . . . .	16
B.3.2 State variables, neutral future reproductive value and its dynamics . . . . .	19
B.3.3 Costate variables: dynamics and interpretation . . . . .	21

B.3.4	Directional selection . . . . .	23
B.3.5	Disruptive selection . . . . .	24
B.3.6	Directional and disruptive selection under continuous age-structure . . . . .	26
B.3.7	Directional selection of phenotypically plastic traits . . . . .	27
<b>C</b>	<b>Selection in subdivided populations</b>	<b>31</b>
C.1	Homogeneous groups . . . . .	31
C.1.1	Life-cycle events . . . . .	32
C.1.2	Fitness . . . . .	32
C.1.3	Directional selection . . . . .	34
C.1.4	Disruptive selection . . . . .	35
C.1.5	Away from playing the field . . . . .	36
C.1.6	Example 1 . . . . .	38
C.1.7	Example 2 . . . . .	41
C.2	Heterogeneity within groups . . . . .	42
C.2.1	Fitness . . . . .	43
C.2.2	Directional selection . . . . .	46
C.2.3	Disruptive selection . . . . .	47



## A Well-mixed and homogeneous populations

### A.1 An example

To illustrate how a typical adaptive dynamics analysis proceeds, we go through a simple example in this section. We consider a well-mixed and homogeneous population in which individuals go through the following life-cycle events: (i) Individuals interact with one another accumulating payoff; (ii) Individuals reproduce, with a fecundity according to payoff and density, leading to frequency- and density-dependence; (iii) Individuals survive or die with a fixed probability and offspring establish to become adults (so that generations are overlapping but we assume there is no effect of age).

The basis of an analysis in a well-mixed and homogeneous population is the individual fitness of a mutant individual, which recall is the expected number of direct descendants produced between two demographic time points, i.e. one full iteration of the cycle (i)-(iii) above. For the scenario outlined in the preceding paragraph, we may write this as,

$$w(z_m, z) = \underbrace{1 - \mu}_{\text{survival}} + \underbrace{\frac{\pi(z_m, z)}{1 + \gamma \hat{n}(z)}}_{\text{fecundity}}, \quad (\text{A-1})$$

where  $1 - \mu$  is the probability of surviving to the next time point (so that  $\mu$  is mortality);  $\pi(z_m, z)$  is the payoff obtained by a mutant in a resident population; and  $\hat{n}(z)$  is the equilibrium density in the resident population (so that  $\gamma$  captures the strength of density-dependence). This demographic equilibrium  $\hat{n}(z)$  must be such that individual fitness in a resident population is one (i.e. such that individuals on average replace themselves),

$$w(z, z) = 1. \quad (\text{A-2})$$

Substituting eq. (A-1) into eq. (A-2) and re-arrangements lead to the equilibrium density

$$\hat{n}(z) = \frac{1}{\gamma} \left( \frac{\pi(z, z)}{\mu} - 1 \right), \quad (\text{A-3})$$

which as expected increases as density-dependence  $\gamma$  weakens, payoff  $\pi(z, z)$  increases, and mortality  $\mu$  decreases. Plugging eq. (A-3) into eq. (A-1) we finally obtain

$$w(z_m, z) = 1 - \mu + \mu \frac{\pi(z_m, z)}{\pi(z, z)}, \quad (\text{A-4})$$

for individual fitness, where we recognise survival in  $1 - \mu$ , and where fecundity can be read as the product between: the proportion  $\mu$  of spots left open by the death of adults; and the probability  $\pi(z_m, z)/\pi(z, z)$  that such a breeding spot is filled by the offspring of a mutant.

For social interactions (event (i) in the life-cycle), let us consider a scenario where individuals randomly pair up

and play a game such that the evolving trait  $0 \leq z \leq 1$  captures some individual investment into cooperation. Individual fecundity can be assumed to read as,

$$\pi(z_m, z) = f_0 \left[ \underbrace{1 - z_m}_{\text{cost}} + \underbrace{B_1(z_m + z) + \frac{B_2}{2}(z_m^2 + z^2) - B_3 z_m z}_{\text{benefit}} \right], \quad (\text{A-5})$$

where  $f_0 > 0$  is fecundity in the absence of cooperation (when  $z_m = z = 0$ );  $1 - z_m$  is the individual cost of cooperation (so that cooperation has a baseline cost of 1); and the rest is the mutual benefit with  $B_1$ ,  $B_2$  and  $B_3$  constants respectively capturing the additive, quadratic and multiplicative effects of cooperation (this part of eq. A-5 can thus be seen as a second order polynomial approximation of a more complex benefit function that depends on  $z_m$  and  $z$ ). The sign of  $B_3$  allows us to consider classical games in the social evolution literature, with for instance  $B_3 > 0$  under the snowdrift game (with antagonistic effects among partners) and  $B_3 < 0$  under the stag-hunt game (with complementary effects among partners).

Substituting eq. (A-5) into eq. (A-4) and deriving according to eq. (3), we obtain the selection gradient,

$$s(z) = B_1 - 1 - z(B_3 - B_2), \quad (\text{A-6})$$

which shows immediately that cooperation increases when absent only if the additive benefit exceeds the cost (i.e.  $s(0) > 0$  only if  $B_1 > 1$ ). Assuming this is true, directional selection favours the evolution of the intermediate singular strategy

$$z^* = \frac{B_1 - 1}{B_3 - B_2} \quad (\text{A-7})$$

(such that  $s(z^*) = 0$ ), provided

$$s'(z^*) = B_2 - B_3 < 0, \quad (\text{A-8})$$

so provided the effects of cooperation are more antagonistic than they are accelerating (so that  $B_2 < B_3$ ). Finally, such strategy is uninvadable when

$$h(z^*) = B_2 < 0, \quad (\text{A-9})$$

in which case the population remains monomorphic for  $z^*$  with equilibrium density,

$$\hat{n}(z^*) = \frac{1}{\gamma} \left[ \frac{f_0}{\mu} (1 + B_1 z^*) - 1 \right]. \quad (\text{A-10})$$

(found by plugging eq. A-7 into eq. A-5 which is in turn plugged into eq. A-3). This equilibrium density increases with the level of cooperation  $z^*$  in the population, as expected.

If by contrast eq. (A-9) does not hold ( $B_2 > 0$ ) and eq. (A-8) does, gradual evolution should lead to the emergence of two differentiated morphs: one that invests more resources into cooperation (“cooperators”) and the other less (“defectors”). These two are then maintained under negative frequency-dependence selection whereby defectors are at an advantage when rare as they can exploit cooperators, but at a disadvantage when common

as they interact with one another and do not reap the benefits of cooperation.

## A.2 Selection in terms of payoff

Here we consider the case where fitness depends on some intermediate payoff (e.g. calories, number of mates, level of light) and show eq. (8) of the main text. We consider the case where individual fitness can be written as a function,

$$w(z_m, z) = w_f(\pi(z_m, z), \pi(z, z)), \quad (\text{A-11})$$

where  $\pi(z_m, z)$  is the payoff obtained by a mutant individual with trait  $z_m$  in a resident population with trait  $z$  and  $\pi(z, z)$  is the payoff to a resident (an example of such a fitness function is eq. A-4). We assume that fitness increases monotonically with payoff, i.e.

$$\frac{\partial w_f(\pi_m, \pi)}{\partial \pi_m} = K > 0. \quad (\text{A-12})$$

The selection gradient can then be unpacked as

$$s(z) = \frac{\partial w_f(\pi(z_m, z), \pi(z, z))}{\partial z_m} = \frac{\partial w_f(\pi_m, \pi)}{\partial \pi_m} \frac{\partial \pi(z_m, z)}{\partial z_m} = K \frac{\partial \pi(z_m, z)}{\partial z_m}, \quad (\text{A-13})$$

as required. Similarly, disruptive selection can be expressed as

$$h(z) = \frac{\partial^2 w_f(\pi(z_m, z), \pi(z, z))}{\partial z_m^2} = \frac{\partial w_f(\pi_m, \pi)}{\partial \pi_m} \frac{\partial^2 \pi(z_m, z)}{\partial z_m^2} + \frac{\partial^2 w_f(\pi_m, \pi)}{\partial \pi_m^2} \left[ \frac{\partial \pi(z_m, z)}{\partial z_m} \right]^2, \quad (\text{A-14})$$

which at a singular strategy  $z^*$  (so that  $\partial \pi(z_m, z) / (\partial z_m) = 0$ ) reduces to

$$h(z^*) = \frac{\partial w_f(\pi_m, \pi)}{\partial \pi_m} \frac{\partial^2 \pi(z_m, z)}{\partial z_m^2} = K \frac{\partial^2 \pi(z_m, z)}{\partial z_m^2}, \quad (\text{A-15})$$

as required.

## A.3 Connection between selection and demography

One other useful aspect of homogeneous and well-mixed populations is that they afford a simple connection between selection and demography. As we saw in Appendix A.1 (eq. A-2), if  $\hat{n}(z)$  is the equilibrium density in a population monomorphic for  $z$ , then such equilibrium is characterised by

$$w(z_m, z, \hat{n}(z)) = 1, \quad (\text{A-16})$$

where we have explicitly written the dependence of fitness on  $\hat{n}(z)$ . Let us assume that at the singular strategy  $z^*$ , this equilibrium condition is satisfied at a point attractor  $\hat{n}(z^*)$  of demography. Differentiating both sides of this eq. (A-16) with respect to  $z$  and some re-arrangements yield that at a singular strategy  $z^*$ , the rate of

change of population size with the trait value is,

$$\hat{n}'(z^*) = \left[ -\frac{\partial w(z_m, z, \hat{n})}{\partial \hat{n}} \right]^{-1} \times \frac{\partial w(z_m, z, \hat{n})}{\partial z} \quad (\text{A-17})$$

[1]. The first factor of this equation can be understood as the effect of density-dependence, where  $\partial w(z_m, z, \hat{n})/(\partial \hat{n}) < 0$  is the effect of an increase in density on individual fitness; the second factor, meanwhile, is the effect of frequency-dependence, with  $\partial w(z_m, z, \hat{n})/(\partial z)$  the effect of a trait change in others on the fitness of a focal individual. Eq. (A-17) reveals that natural selection leads to a demographic maximum (where  $\hat{n}'(z^*) = 0$ ) only in the absence of frequency-dependence (i.e. where  $\partial w(z_m, z, \hat{n})/(\partial z) = 0$ ). Otherwise, population size would be greater if prosocial traits (such that  $\partial w(z_m, z, \hat{n})/(\partial z) > 0$ ) were more greatly expressed than at their singular value, and conversely, if antisocial traits (such that  $\partial w(z_m, z, \hat{n})/(\partial z) < 0$ ) were lesser expressed, especially so where density-dependence is weak (where  $\partial w(z_m, z, \hat{n})/(\partial z)$  is close to zero). More generally, eq. (A-17) highlights how selection leads to an inefficient outcome at the population level as soon as there is frequency-dependence.

## B Selection in class-structured populations

### B.1 Matrix population models

We first consider a population that is divided into a finite number  $M$  of discrete classes, e.g. males and females, juvenile and mature individuals, or subordinate and dominant individuals, deriving eqs. (10)-(12) of the main text. Our derivations largely owe to [2].

#### B.1.1 Invasion fitness

The joint dynamics of the number of mutants across classes can be modelled by a matrix equation,

$$\mathbf{n}_{t+1} = \mathbf{W}(z_m, z) \cdot \mathbf{n}_t \quad (\text{B-18})$$

where entry  $i \in \{1, \dots, M\}$  of the vector  $\mathbf{n}_t$  gives the number of mutants in class  $i$  at some time  $t$ , and the  $(i, j)$ -entry of the  $M \times M$  matrix,  $\mathbf{W}(z_m, z)$ , which we denote by  $w_{ij}(z_m, z)$ , is the expected number of mutants in class  $i$  produced by a mutant in class  $j$  (when the mutant is rare and the resident population is at equilibrium). The matrix  $\mathbf{W}(z_m, z)$  is sometimes referred to as the mean matrix in the theory of multi-type branching process. From this theory, we know that invasion fitness is given by the leading eigenvalue  $\rho(z_m, z)$  of matrix  $\mathbf{W}(z_m, z)$ . As such,  $\rho(z_m, z)$  satisfies,

$$\rho(z_m, z) \mathbf{q}(z_m, z) = \mathbf{W}(z_m, z) \cdot \mathbf{q}(z_m, z), \quad (\text{B-19})$$

where  $\mathbf{q}(z_m, z)$  is the right eigenvector of  $\mathbf{W}(z_m, z)$ . We scale this vector such that its entries sum to one, i.e. such that

$$\mathbf{q}(z_m, z) \cdot (1, 1, \dots, 1) = 1. \quad (\text{B-20})$$

In this case the  $i$ -entry of  $\mathbf{q}(z_m, z)$  corresponds to the asymptotic probability that a randomly sampled mutant is in an individual in class  $i$ .

**Reproductive value.** Next, let us define  $\mathbf{v}^\circ$  as the left eigenvector of the mean matrix  $\mathbf{W}^\circ = \mathbf{W}(z, z)$  under neutrality (whose  $(i, j)$ -entry gives the expected number of individuals in class  $i$  produced by an individual in class  $j$  in the resident population at equilibrium). Throughout, we denote quantities under neutrality, i.e. when  $z_m = z$ , by a superscript  $\circ$ . Such quantities should thus be read as functions of the resident trait  $z$  (e.g.  $\mathbf{v}^\circ$  is a function of  $z$ ) but we do not write such dependency explicitly to avoid notational clutter. Using the property of eigenvectors and the fact that invasion fitness of a neutral mutant is one (i.e. the eigenvalue associated to  $\mathbf{v}^\circ$  is 1), we have,

$$\mathbf{v}^\circ \cdot \mathbf{W}^\circ = \mathbf{v}^\circ. \quad (\text{B-21})$$

It will turn out to be useful to normalise  $\mathbf{v}^\circ$  such that

$$\mathbf{v}^\circ \cdot \mathbf{q}^\circ = 1, \quad (\text{B-22})$$

where  $\mathbf{q}^\circ$  is the right eigenvector of the neutral mean matrix  $\mathbf{W}^\circ$  (such that  $\mathbf{W}^\circ \cdot \mathbf{q}^\circ = \mathbf{q}^\circ$ ). The  $i$ -entry of the left eigenvector  $\mathbf{v}^\circ$  can then thought of as the “normalised reproductive value” of an individual in class  $i$ , i.e. its relative (compared to other classes) asymptotic demographic contribution to the future of the population in the absence of selection, normalised such that the average reproductive value is one (see e.g. p. 97 of [3]).

**Weighted fitness effects.** The interpretation of  $q_j(z_m, z)$  as the asymptotic probability that a randomly sampled mutant is in an individual in class  $i$ , and of  $v_i^\circ$  as the “reproductive value” of an individual in class  $i$  can then help understand selection in class-structured populations. Indeed, left-multiplying both sides of eq. (B-19) by  $\mathbf{v}^\circ$ , we obtain after some re-arrangements,

$$\rho(z_m, z) = \frac{\mathbf{v}^\circ \cdot \mathbf{W}(z_m, z) \cdot \mathbf{q}(z_m, z)}{\mathbf{v}^\circ \cdot \mathbf{q}(z_m, z)} = \frac{1}{V(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ w_{ij}(z_m, z) q_j(z_m, z) \quad (\text{B-23})$$

where we have defined

$$V(z_m, z) = \mathbf{v}^\circ \cdot \mathbf{q}(z_m, z) = \sum_{i=1}^M v_i^\circ q_i(z_m, z). \quad (\text{B-24})$$

To capture the effects of selection, we denote by

$$\alpha_{ij}(z_m, z) = w_{ij}(z_m, z) - w_{ij}^\circ \quad (\text{B-25})$$

the difference in the expected number of offspring produced by a mutant and resident (according to class). Substituting eq. (B-25) into eq. (B-23), we can re-arrange invasion fitness to read as,

$$\begin{aligned}
\rho(z_m, z) &= \frac{1}{V(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \left[ w_{ij}^\circ + \alpha_{ij}(z_m, z) \right] q_j(z_m, z) \\
&= \frac{1}{V(z_m, z)} \underbrace{\sum_{j=1}^M \sum_{i=1}^M v_i^\circ w_{ij}^\circ q_j(z_m, z)}_{=V(z_m, z)} + \frac{1}{V(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \alpha_{ij}(z_m, z) q_j(z_m, z) \\
&= 1 + \frac{1}{V(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \alpha_{ij}(z_m, z) q_j(z_m, z).
\end{aligned} \tag{B-26}$$

This equation shows that  $\rho(z_m, z) > 1$  if and only if the second term,  $\sum_{i=1}^M \sum_{j=1}^M v_i^\circ \alpha_{ij}(z_m, z) q_j(z_m, z)$ , is positive. It is the sum across classes  $j$  of the probability that a mutant is in class  $j$  ( $q_j(z_m, z)$ ) times the number of mutants in class  $i$  produced by such a mutant in excess to a resident ( $\alpha_{ij}(z_m, z)$ ), where these descendants are weighted by their reproductive value in the resident population ( $v_i^\circ$ ). Note that the choice of the left eigenvector  $v^\circ$  as the vector of reproductive values ensures that any neutral mutant ( $z_m = z$ ) has invasion fitness equal to one. As such, this vector is independent from any mutant effect.

### B.1.2 Directional selection

Substituting eq. (B-26) into eq. (3), we obtain that the selection gradient can be expressed as,

$$\begin{aligned}
s(z) &= \underbrace{\frac{1}{V^\circ}}_{=1} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial}{\partial z_m} [\alpha_{ij}(z_m, z) q_j(z_m, z)] + \frac{\partial}{\partial z_m} \left[ \frac{1}{V(z_m, z)} \right] \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \underbrace{\alpha_{ij}^\circ}_{=0} q_j^\circ \\
&= \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial \alpha_{ij}(z_m, z)}{\partial z_m} q_j^\circ + \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \underbrace{\alpha_{ij}^\circ}_{=0} \frac{\partial q_j(z_m, z)}{\partial z_m} \\
&= \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial \alpha_{ij}(z_m, z)}{\partial z_m} q_j^\circ
\end{aligned} \tag{B-27}$$

where here and hereafter all derivatives are evaluated at  $z_m = z$ . Since the derivatives of  $\alpha_{ij}(z_m, z)$  and  $w_{ij}(z_m, z)$  with respect to  $z_m$  are equal at all orders (from eq. B-25), i.e. since

$$\frac{\partial^a \alpha_{ij}(z_m, z)}{\partial z_m^a} = \frac{\partial^a w_{ij}(z_m, z)}{\partial z_m^a}, \tag{B-28}$$

we finally have

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial w_{ij}(z_m, z)}{\partial z_m} q_j^\circ \tag{B-29}$$

for the selection gradient. See eq. (10) in main text for interpretation.

### B.1.3 Disruptive selection

Similarly, plugging eq. (B-26) into eq. (3), we have at a singular strategy that disruptive selection is given by

$$\begin{aligned}
h(z^*) &= \underbrace{\frac{1}{V^\circ} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial^2}{\partial z_m^2} [\alpha_{ij}(z_m, z) q_j(z_m, z)]}_{=1} + \frac{\partial^2}{\partial z_m^2} \left[ \frac{1}{V(z_m, z)} \right] \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \underbrace{\alpha_{ij}^\circ q_j^\circ}_{=0} \\
&\quad + 2 \frac{\partial}{\partial z_m} \left[ \frac{1}{V(z_m, z)} \right] \underbrace{\sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial}{\partial z_m} [\alpha_{ij}(z_m, z) q_j(z_m, z)]}_{=0, \text{ when } z_m = z = z^*} \\
&= \sum_{i=1}^M \sum_{j=1}^M \left[ v_i^\circ \frac{\partial^2 \alpha_{ij}(z_m, z)}{\partial z_m^2} q_j^\circ + 2 v_i^\circ \frac{\partial \alpha_{ij}(z_m, z)}{\partial z_m} \frac{\partial q_j(z_m, z)}{\partial z_m} + v_i^\circ \underbrace{\alpha_{ij}^\circ}_{=0} \frac{\partial^2 q_j(z_m, z)}{\partial z_m^2} \right].
\end{aligned} \tag{B-30}$$

Using eq. (B-28), we finally obtain

$$h(z^*) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial^2 w_{ij}(z_m, z)}{\partial z_m^2} q_j^\circ + 2 \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \frac{\partial w_{ij}(z_m, z)}{\partial z_m} \frac{\partial q_j(z_m, z)}{\partial z_m}. \tag{B-31}$$

See eqs. (11)-(12) in main text for interpretation.

## B.2 Age-structure

The case of age-structure allows for further analysis and many connections to previous literature, especially on life history evolution. Here we derive eqs. (14) and (15) of the main text.

### B.2.1 Leslie matrix

The nature of age gives the mean matrix a special structure:

$$W(z_m, z) = \begin{pmatrix} b_1(z_m, z) & b_2(z_m, z) & \dots & b_{M-1}(z_m, z) & b_M(z_m, z) \\ p_1(z_m, z) & 0 & \dots & 0 & 0 \\ 0 & p_2(z_m, z) & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & p_{M-1}(z_m, z) & 0 \end{pmatrix}, \tag{B-32}$$

i.e.,

$$w_{ij}(z_m, z) = \begin{cases} b_j(z_m, z), & i = 1 \\ p_j(z_m, z), & i = j + 1 \\ 0, & \text{otherwise} \end{cases} \tag{B-33}$$

where  $b_j(z_m, z)$  is the expected number of offspring that a mutant individual of age  $j$  produces (and thus end up in age class 1) and  $p_j(z_m, z)$  is the probability that a mutant individual of age  $j$  survives to age  $j + 1$ . The matrix eq. (B-32) is often referred to as a Leslie matrix, very commonly used in demography research. The special structure of this matrix, with its many zero entries, allows simplifications as we review below.

## B.2.2 Stable age distribution and Fisher's reproductive value

We can first revisit the vectors  $\mathbf{q}(z_m, z)$  and  $\mathbf{v}^\circ$  that are necessary to characterise selection. Using standard results [e.g. 4, p. 87], the normalised right eigenvector  $\mathbf{q}(z_m, z)$  of the Leslie matrix eq. (B-32) has entry  $j$  given by,

$$q_j(z_m, z) = \frac{l_j(z_m, z)}{\rho(z_m, z)^{j-1}} \times \left[ \sum_{k=1}^M \frac{l_k(z_m, z)}{\rho(z_m, z)^{k-1}} \right]^{-1}, \quad (\text{B-34})$$

where

$$l_j(z_m, z) = \prod_{k=1}^{j-1} p_k(z_m, z) \quad (\text{B-35})$$

is the probability that a mutant individual survives at least until age  $j$ . Accordingly,  $q_j(z_m, z)$  gives the asymptotic mutant age-distribution (i.e.  $q_j(z_m, z)$  is the asymptotic probability that a randomly sampled mutant is of age  $j$ ). In a population monomorphic for the resident  $z$ , eq. (B-34) reveals this probability reduces to

$$q_j^\circ = \frac{l_j^\circ}{L^\circ} \quad (\text{B-36})$$

(since  $\rho(z, z) = 1$ ), where

$$L^\circ = \sum_{j=1}^M l_j^\circ, \quad (\text{B-37})$$

is the expected lifespan in a population monomorphic for  $z$ .

The left eigenvector that is relevant to our analysis,  $\mathbf{v}^\circ$ , is found by solving eq. (B-21) (with  $\mathbf{W}^\circ$  given by eq. B-32) subject to the constraint eq. (B-22) (with  $\mathbf{q}^\circ$  given by eq. B-36). Building on Fisher's seminal work [5], the literature on life-history evolution often uses a different constraint for this left eigenvector, which we denote by  $\tilde{\mathbf{v}}^\circ$  and refer to as "current reproductive value" (to contrast with "normalised reproductive value",  $\mathbf{v}^\circ$ ). Instead of eq. (B-22),  $\tilde{\mathbf{v}}^\circ$  is characterised by the constraint that

$$\tilde{v}_1^\circ = 1, \quad (\text{B-38})$$

meaning that the current reproductive value of an individual in age class 1 is one [4, 6]. Relative to this, the current reproductive value of an individual of age  $j$  is then given by,

$$\tilde{v}_j^\circ = \sum_{k=j}^M \frac{l_k^\circ b_k^\circ}{l_j^\circ} \quad (\text{B-39})$$

(e.g. eq. 4.89 in [4]). Eq. (B-39) is easy to interpret:  $\tilde{v}_j^\circ$  is the expected number of offspring that an individual



produces given it has survived to age  $j$  (under neutrality).

The connection between  $\tilde{\mathbf{v}}^\circ$  and  $\mathbf{v}^\circ$  can be seen by noting that since both are eigenvectors of the same matrix, they must be proportional, i.e.,

$$\mathbf{v}^\circ = C \tilde{\mathbf{v}}^\circ, \quad (\text{B-40})$$

where  $C$  is a constant. Plugging eq. (B-40) into eq. (B-22), we obtain,

$$\sum_{j=1}^M v_j^\circ q_j^\circ = 1 \iff C \sum_{j=1}^M \tilde{v}_j^\circ q_j^\circ = 1 \iff C \sum_{j=1}^M \sum_{k=j}^M \frac{l_k^\circ b_k^\circ}{l_j^\circ} \frac{l_j^\circ}{L^\circ} = 1 \quad (\text{B-41})$$

where we used eqs. (B-39) and (B-36). This simplifies to

$$\frac{C}{L^\circ} \underbrace{\sum_{j=1}^M \sum_{k=j}^M l_k^\circ b_k^\circ}_{=\sum_{j=1}^M j l_j^\circ b_j^\circ = T^\circ} = 1, \quad (\text{B-42})$$

where  $T^\circ$  is the generation time in a population monomorphic for  $z$ , i.e. the expected age of a parent. Rearranging eq. (B-42) to solve for  $C$  and substituting the result into eq. (B-40), we obtain

$$\mathbf{v}^\circ = \frac{L^\circ}{T^\circ} \tilde{\mathbf{v}}^\circ \quad (\text{B-43})$$

(in line with eqs. 1.54 and 1.56 in [6]). Using eq. (B-38), we then obtain from eq. (B-43) that

$$v_1^\circ = \frac{L^\circ}{T^\circ}. \quad (\text{B-44})$$

The reproductive value of an individual of age 1 is thus given by the ratio of lifespan to generation time, which can be viewed as a measure of generational overlap. Further, it follows from eqs. (B-43) and (B-22) that

$$\tilde{\mathbf{v}}^\circ \cdot \mathbf{q}^\circ = \frac{T^\circ}{L^\circ}, \quad (\text{B-45})$$

i.e. the average current reproductive value is equal to  $T^\circ/L^\circ$ .

### B.2.3 Directional selection in age-structured populations

We can then use the above to characterise directional and disruptive selection in age-structured populations.

First, plugging eq. (B-33) into eq. (B-29), we obtain that the selection gradient can be expressed as,

$$s(z) = \sum_{j=1}^M \left[ v_1^\circ \frac{\partial b_j(z_m, z)}{\partial z_m} + v_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] q_j^\circ. \quad (\text{B-46})$$

Factoring by  $v_1^\circ$  and using eq. (B-36) for  $q_j^\circ$ , this selection gradient can alternatively be expressed as

$$\begin{aligned} s(z) &= \frac{v_1^\circ}{L^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \frac{v_{j+1}^\circ}{v_1^\circ} \frac{\partial p_j(z_m, z)}{\partial z_m} \right] l_j^\circ \\ &= \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \tilde{v}_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] l_j^\circ \end{aligned} \quad (\text{B-47})$$

where we used eqs. (B-43), (B-39) and (B-44) to go from the first to second line. We can define

$$\mu_j(z_m, z) = 1 - p_j(z_m, z) \quad (\text{B-48})$$

as the probability of death at age  $j$  for a mutant  $z_m$  and thus rewrite the selection gradient as

$$s(z) = \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} - \tilde{v}_{j+1}^\circ \frac{\partial \mu_j(z_m, z)}{\partial z_m} \right] l_j^\circ. \quad (\text{B-49})$$

See eq. (14) in main text for interpretation.

#### B.2.4 Disruptive selection in age-structured populations

Similarly, substituting eq. (B-33) into eq. (B-31), we find that disruptive selection at a singular strategy  $z^*$  can be decomposed as the sum of two terms,

$$h(z^*) = h_w(z^*) + 2h_q(z^*) \quad (\text{B-50})$$

where

$$h_w(z^*) = \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial^2 b_j(z_m, z)}{\partial z_m^2} + \tilde{v}_{j+1}^\circ \frac{\partial^2 p_j(z_m, z)}{\partial z_m^2} \right] l_j^\circ \quad (\text{B-51})$$

and

$$h_q(z^*) = \frac{L^\circ}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \tilde{v}_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] \frac{\partial q_j(z_m, z)}{\partial z_m}. \quad (\text{B-52})$$

This latter term can be made more simple by first noting that at a singular strategy, where by definition,

$$\left. \frac{\partial \rho(z_m, z)}{\partial z_m} \right|_{z_m=z=z^*} = 0, \quad (\text{B-53})$$

the derivative of the age distribution (from eq. B-34) reduces to,

$$\begin{aligned} \left. \frac{\partial q_j(z_m, z)}{\partial z_m} \right|_{z_m=z=z^*} &= \frac{\partial}{\partial z_m} \left[ \frac{l_j(z_m, z)}{\sum_{k=1}^M l_k(z_m, z)} \right]_{z_m=z=z^*} \\ &= \frac{1}{L^{\circ 2}} \left( \frac{\partial l_j(z_m, z)}{\partial z_m} L^\circ + l_j^\circ \frac{\partial L(z_m, z)}{\partial z_m} \right) \end{aligned} \quad (\text{B-54})$$

where

$$L(z_m, z) = \sum_{k=1}^M l_k(z_m, z) \quad (\text{B-55})$$

is the expected lifespan of a mutant  $z_m$  in a resident population  $z$ . Substituting eq. (B-54) into eq. (B-52) then yields

$$\begin{aligned} h_q(z^*) &= \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \tilde{v}_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] \frac{\partial l_j(z_m, z)}{\partial z_m} \\ &+ \frac{1}{T^\circ L^\circ} \underbrace{\sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \tilde{v}_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] l_j^\circ}_{\propto s(z)=0 \text{ when } z=z^*} \times \frac{\partial L(z_m, z)}{\partial z_m}, \end{aligned} \quad (\text{B-56})$$

leaving us with

$$h_q(z^*) = \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} + \tilde{v}_{j+1}^\circ \frac{\partial p_j(z_m, z)}{\partial z_m} \right] \frac{\partial l_j(z_m, z)}{\partial z_m}. \quad (\text{B-57})$$

Finally, using eq. (B-48), the two components of disruptive selection can be expressed as

$$\begin{aligned} h_w(z^*) &= \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial^2 b_j(z_m, z)}{\partial z_m^2} - \tilde{v}_{j+1}^\circ \frac{\partial^2 \mu_j(z_m, z)}{\partial z_m^2} \right] l_j^\circ \\ h_q(z^*) &= \frac{1}{T^\circ} \sum_{j=1}^M \left[ \frac{\partial b_j(z_m, z)}{\partial z_m} - \tilde{v}_{j+1}^\circ \frac{\partial \mu_j(z_m, z)}{\partial z_m} \right] \frac{\partial l_j(z_m, z)}{\partial z_m} \end{aligned} \quad (\text{B-58})$$

See main text eq. (15) for interpretation.

We can connect eqs. (B-49) and (B-58) with marginal effects on the reproductive number,

$$R_0(z_m, z) = \sum_{j=1}^M l_j(z_m, z) b_j(z_m, z), \quad (\text{B-59})$$

which is the expected number of offspring produced by a mutant individual over its whole lifetime. This reproductive number is a classical measure of reproductive success, which is sometimes more convenient to work with than  $\rho(z_m, z)$ . In fact, using the next generation theorem, it is straightforward to show that  $\rho(z_m, z) > 1$  if and only if  $R_0(z_m, z) > 1$ . We can quantify this connection by using the fact that due to the structure of the Leslie matrix, its leading eigenvalue  $\rho(z_m, z)$  satisfies the so-called Euler-Lotka equation,

$$\sum_{j=1}^M \frac{l_j(z_m, z) b_j(z_m, z)}{\rho(z_m, z)^j} = 1 \quad (\text{B-60})$$

Deriving both sides of this equation with respect to  $z_m$  and estimating it at  $z$ , we obtain

$$\frac{\partial}{\partial z_m} \underbrace{\left[ \sum_{j=1}^M l_j(z_m, z) b_j(z_m, z) \right]}_{=R_0(z_m, z)} - \underbrace{\frac{\partial \rho(z_m, z)}{\partial z_m}}_{=s(z)} \underbrace{\sum_{j=1}^M j l_j^\circ b_j^\circ}_{=T^\circ} = 0. \quad (\text{B-61})$$

Re-arranging the above yields

$$s(z) = \frac{1}{T^\circ} \frac{\partial R_0(z_m, z)}{\partial z_m}. \quad (\text{B-62})$$

Similarly, deriving eq. (B-60) twice with respect to  $z_m$  and estimating at a singular strategy  $z_m = z = z^*$ , we obtain

$$h(z^*) = \frac{1}{T^\circ} \left. \frac{\partial^2 R_0(z_m, z)}{\partial z_m^2} \right|_{z_m=z=z^*}. \quad (\text{B-63})$$

Hence directional selection  $s(z)$ , and disruptive selection  $h(z^*)$  (at a singular strategy), are proportional to the marginal effects of trait expression on  $R_0(z_m, z)$ , with the constant of proportionality the inverse of neutral generation time.

### B.2.5 Example

In this section we go through the analysis of an example of trait evolution in an aged-structured population. The main aim is to illustrate how the second term  $h_q(z^*)$  of disruptive selection can lead to polymorphism (so we do not perform an exhaustive analysis of this example). We consider a simple scenario where there are just two age classes  $M = 2$ . The evolving trait  $0 \leq z \leq 1$  is the proportion of resources invested into reproduction at age 1, such that the expected number of offspring produced by a focal mutant individual with trait  $z_m$  is

$$b_1(z_m, z) = \frac{f_1 z_m}{1 + \gamma_1 \hat{n}_1(z)}, \quad (\text{B-64})$$

where  $f_1 > 0$  is a parameter that determines the conversion of resources into offspring at age 1;  $\hat{n}_1(z)$  is the equilibrium density of individuals of age 1 in the resident population; and  $\gamma_1 > 0$  is a parameter for the strength of density-dependent competition among individuals of age 1 (we assume there is no density-dependent competition among individuals of different ages for simplicity, our results are qualitatively similar as long as density-dependent competition among individuals of different ages is weaker than among of the same age).

Investing resources into reproduction at age 1 however diverts from other vital functions so that survival from age 1 to 2 is impaired according to

$$l_2(z_m, z) = p_1(z_m, z) = 1 - z_m; \quad (\text{B-65})$$

and fecundity at age 2 decreases with  $z_m$  according to,

$$b_2(z_m, z) = \frac{f_2 (1 - z_m)^\beta}{1 + \gamma_2 \hat{n}_2(z)}, \quad (\text{B-66})$$

where  $\beta > 0$  modulates the effect of fewer resources available on fecundity;  $f_2 > 0$  determines the conversion of resources into offspring at age 2;  $\gamma_2 > 0$  controls the strength of density-dependent competition among individuals of age 2; and

$$\hat{n}_2(z) = p_1^\circ \hat{n}_1(z) = (1 - z) \hat{n}_1(z) \quad (\text{B-67})$$

is the equilibrium density of individuals of age 2 in the resident population, which depends on  $\hat{n}_1(z)$ . This demographic equilibrium is determined from the fact that in a monomorphic at equilibrium, the expected number of offspring produced during one's lifetime is one, i.e. from the fact that

$$R_0^\circ = \sum_{j=1}^M l_j^\circ b_j^\circ = 1, \quad (\text{B-68})$$

which for our model here is,

$$\frac{f_1 z}{1 + \gamma_1 \hat{n}_1(z)} + \frac{f_2 (1-z)^{1+\beta}}{1 + \gamma_2 (1-z) \hat{n}_1(z)} = 1. \quad (\text{B-69})$$

Eq. (B-69) can be solved analytically for  $\hat{n}_1(z)$  but we do not present its solution here as it is complicated and not particularly illuminating.

The above gives all the necessary components to compute the selection gradient (eq. B-49), from which we find that a singular strategy  $z^*$  must be such that

$$\frac{f_1}{1 + \gamma_1 \hat{n}_1(z^*)} = \frac{f_2 (1-z^*)^\beta (1+\beta)}{1 + (1-z^*) \hat{n}_1(z^*)}, \quad (\text{B-70})$$

where the left hand side is the marginal benefit from an increase in the investment in fecundity at age 1 and the right hand side is the marginal cost. We can solve for this singular value numerically and focus on the case where it is convergence stable, which a numerical inspection suggests occurs where  $\beta < 1$ . From the above, we can also quantify disruptive selection (eq. B-50 and eq. B-58), which at a singular strategy we find can be expressed as

$$h(z^*) = \underbrace{\frac{1}{T^\circ} \frac{f_2 \beta (\beta-1) (1-z^*)^{\beta-1}}{1 + \gamma_2 (1-z^*) \hat{n}_1(z^*)}}_{=h_w(z^*) < 0 \text{ when } \beta < 1} + 2 \underbrace{\frac{1}{T^\circ} \frac{f_2 \beta (1-z^*)^{\beta-1}}{1 + \gamma_2 (1-z^*) \hat{n}_1(z^*)}}_{=h_q(z^*) > 0} = \frac{f_2 \beta (1+\beta) (1-z^*)^{\beta-1}}{1 + \gamma_2 (1-z^*) \hat{n}_1(z^*)} > 0. \quad (\text{B-71})$$

Eq. (B-71) reveals that selection is always disruptive in this model and that this is due to the  $h_q(z^*)$  term. In other words, selection is disruptive because an increased investment in fecundity at age 1 has antagonistic pleiotropic effects on both the probability of surviving till age 2 ( $\partial l_2(z_m, z)/(\partial z_m) < 0$ ) and fecundity at age 2 ( $\partial b_2(z_m, z)/(\partial z_m) < 0$ ). This allows for two morphs to emerge: (i) one that expresses large  $z$  and invests most of its resources into fecundity at age 1 at the expense of age 2; and (2) one expresses small  $z$  and does not reproduce at age 1 to ensure it can survive to age 2 and reproduce.

### B.3 Physiological structure

Here we consider the case where individuals are in different physiological states that vary with age according to eq. (16) of the main text, and derive eqs. (19), (20) and (II.A) (as well as eq. (I.A) for plastic traits). To connect more straightforwardly to existing literature and methods, age is assumed to be continuous in this model.

### B.3.1 Selection under continuous age and physiological structure

We first characterise selection under continuous age-structure. Derivations here follow a line of logic similar to Appendix A of [7]. In our model, there is a direct correspondence between physiological state (state for short) and age (i.e. for each age, the state is known, specifically given by eq. 16). In this case, invasion fitness satisfies the continuous version of the Euler-Lotka equation (continuous version of eq. B-60). Letting

$$\mathbf{z} = (z_m, z) \quad (\text{B-72})$$

collect resident and mutant traits for short, we thus have,

$$\int_0^M \exp(-a\rho(\mathbf{z})) b(\mathbf{z}, x(a)) l(a) da = 1, \quad (\text{B-73})$$

where  $\rho(\mathbf{z})$  is the invasion fitness of the mutant,  $b(\mathbf{z}, x(a))$  is the fecundity rate of a mutant individual in state  $x(a)$  at age  $a$ ,  $l(a)$  is the probability that a mutant individual survives to age  $a$ . The dynamical equations that characterise  $x(a)$  and  $l(a)$  are given by eqs. (16) and (17) of the main text, respectively. We assume that there is *a priori* no maximum age,  $M \rightarrow \infty$ , so that lifespan is endogenously determined by the evolving trait  $z$ . Since time is continuous, the invasion fitness of a neutral mutant is zero (instead of one in discrete time) as it is measured on an exponential scale, i.e.

$$\rho(\mathbf{z}^\circ) = 0 \quad (\text{B-74})$$

where  $\mathbf{z}^\circ = (z, z)$ .

Let us define

$$\phi(\rho(\mathbf{z})) = \int_0^M \exp(-a\rho(\mathbf{z})) b(\mathbf{z}, x(a)) l(a) da. \quad (\text{B-75})$$

We can then express eq. (B-73) as

$$\phi(\rho(\mathbf{z})) = 1. \quad (\text{B-76})$$

Recall that  $\epsilon = z_m - z$  is the difference between mutant and resident traits, which is assumed to be small. We then Taylor expand the left hand side of eq. (B-76) around  $\epsilon = 0$  to obtain

$$\phi(\rho(\mathbf{z}^\circ)) + \epsilon \left. \frac{d\phi(\rho(\mathbf{z}))}{d\epsilon} \right|_{\epsilon=0} + \frac{\epsilon^2}{2} \left. \frac{d^2\phi(\rho(\mathbf{z}))}{d\epsilon^2} \right|_{\epsilon=0} + \mathcal{O}(\epsilon^3) = 1. \quad (\text{B-77})$$

Since  $\phi(\rho(\mathbf{z}^\circ)) = 1$  (substitute eq. B-74 into eq. B-75), eq. (B-77) becomes

$$\epsilon \left. \frac{d\phi(\rho(\mathbf{z}))}{d\epsilon} \right|_{\epsilon=0} + \frac{\epsilon^2}{2} \left. \frac{d^2\phi(\rho(\mathbf{z}))}{d\epsilon^2} \right|_{\epsilon=0} + \mathcal{O}(\epsilon^3) = 0. \quad (\text{B-78})$$

This equality must be true for all  $\epsilon$  so that two equations must be satisfied,

$$\left. \frac{d\phi(\rho(\mathbf{z}))}{d\epsilon} \right|_{\epsilon=0} = 0 \quad (\text{B-79a})$$

$$\left. \frac{d^2 \phi(\rho(\mathbf{z}))}{d\epsilon^2} \right|_{\epsilon=0} = 0 \quad (\text{B-79b})$$

which we can use to characterise directional and disruptive selection from invasion fitness, respectively.

**Directional selection.** The selection gradient with the notation of this section is given by

$$s(\mathbf{z}) = \left. \frac{d\rho(\mathbf{z})}{d\epsilon} \right|_{\epsilon=0}. \quad (\text{B-80})$$

To get to this gradient, we unpack the derivative on the left hand side of eq. (B-79a) using the chain and product rules on eq. (B-75), obtaining

$$\left. \frac{d\phi(\rho(\mathbf{z}))}{d\rho(\mathbf{z})} \right|_z \times s(\mathbf{z}) + \int_0^M \left. \frac{\partial b(\mathbf{z}, x(a))}{\partial z_m} \right|_z l^\circ(a) da + \int_0^M b(\mathbf{z}^\circ, x^\circ(a)) \left. \frac{\partial l(a)}{\partial z_m} \right|_z da = 0, \quad (\text{B-81})$$

where we used eq. (B-74) and throughout Appendix B.3, the subscript  $z$  denotes expressions evaluated at the resident strategy (i.e.  $z_m = z$ ). Solving this equation for  $s(\mathbf{z})$  then yields,

$$s(\mathbf{z}) = \frac{1}{-\left. \frac{d\phi(\rho(\mathbf{z}))}{d\rho(\mathbf{z})} \right|_z} \left[ \int_0^M \left( \frac{\partial b(\mathbf{z}, x(a))}{\partial z_m} l^\circ(a) + b(\mathbf{z}^\circ, x^\circ(a)) \frac{\partial l(a)}{\partial z_m} \right) da \right]_z. \quad (\text{B-82})$$

Note however from eq. (B-75) that

$$\left. \frac{d\phi(\rho(\mathbf{z}))}{d\rho(\mathbf{z})} \right|_z = - \int_0^M a b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) da = -T^\circ, \quad (\text{B-83})$$

where  $T^\circ$  is the expected age of a parent, i.e. generation time, under neutrality. We thus have

$$s(\mathbf{z}) = \frac{1}{T^\circ} \left[ \int_0^M \left( \frac{\partial b(\mathbf{z}, x(a))}{\partial z_m} l^\circ(a) + b(\mathbf{z}^\circ, x^\circ(a)) \frac{\partial l(a)}{\partial z_m} \right) da \right]_z, \quad (\text{B-84})$$

where  $\partial b(\mathbf{z}, x(a))/(\partial z_m)$  is the mutant effect on fecundity and  $\partial l(a)/(\partial z_m)$  on survival. These mutant effects on fecundity and survival can be direct or indirect via a change in state  $x(a)$  (i.e. the trait  $z_m$  influences  $x(a)$  according to eq. (16) which in turn influences fecundity and survival). Such indirect effects are not straightforwardly characterised as one must take into account that a change in state at a certain age has cascading fitness effects later on in life. One efficient method to solve this problem comes from optimal control theory. Before using this method in section B.3.4, let us first describe disruptive selection.

**Disruptive selection.** With the above notation, disruptive selection is given by

$$h(\mathbf{z}^*) = \left. \frac{d^2 \rho(\mathbf{z})}{d\epsilon^2} \right|_{\epsilon=0, \mathbf{z}^*}, \quad (\text{B-85})$$

where throughout Appendix B.3, the subscript  $z^*$  denotes expressions evaluated at the singular strategy (i.e. where  $z_m = z = z^*$ ), which recall is such that

$$s(z^*) = 0. \quad (\text{B-86})$$

To obtain disruptive selection, we use the chain and product rules on the left hand side of eq. (B-79b) (with eq. B-75), which estimated at the singular strategy leads us to,

$$\left. \frac{d\phi(\rho(z))}{d\rho(z)} \right|_{z^*} \times h(z^*) + \int_0^M \left( \frac{\partial^2 b(z, x(a))}{\partial z_m^2} l^\circ(a) + 2 \frac{\partial b(z, x(a))}{\partial z_m} \frac{\partial l(a)}{\partial z_m} + b(z^\circ, x^\circ(a)) \frac{\partial^2 l(a)}{\partial z_m^2} \right)_{z^*} da = 0. \quad (\text{B-87})$$

Solving the above for  $h(z^*)$  and using eq. (B-83) then gives,

$$h(z^*) = \frac{1}{T^*} \int_0^M \left( \frac{\partial^2 b(z, x(a))}{\partial z_m^2} l^\circ(a) + 2 \frac{\partial b(z, x(a))}{\partial z_m} \frac{\partial l(a)}{\partial z_m} + b(z^\circ, x^\circ(a)) \frac{\partial^2 l(a)}{\partial z_m^2} \right)_{z^*} da, \quad (\text{B-88})$$

where  $T^* = T^\circ|_{z^*}$  denotes generation time in a population monomorphic for the singular strategy. As eq. (B-84), eq. (B-88) does not show explicitly how selection acts via changes in state  $x(a)$ . We reveal these effects in section B.3.5 but first connect eqs. (B-84) and (B-88) to the basic reproductive number, which will be a useful platform to derive selection.

**Selection in terms of the basic reproductive number.** In an age-structured population, the basic reproductive number is the expected number of offspring produced by an individual over its lifetime. For our model, this is

$$R_0(z) = \int_0^M b(z, x(a)) l(a) da. \quad (\text{B-89})$$

Taking the derivative of the above with respect to  $\epsilon$  then reads as,

$$\left. \frac{dR_0(z)}{d\epsilon} \right|_{\epsilon=0} = \int_0^M \left( \frac{\partial b(z, x(a))}{\partial z_m} l^\circ(a) + b(z^\circ, x^\circ(a)) \frac{\partial l(a)}{\partial z_m} \right)_z da, \quad (\text{B-90})$$

which compared with eq. (B-84) reveals that

$$s(z) = \frac{1}{T^\circ} \left. \frac{dR_0(z)}{d\epsilon} \right|_{\epsilon=0} = \frac{1}{T^\circ} \left. \frac{\partial R_0(z)}{\partial z_m} \right|_z, \quad (\text{B-91})$$

as expected (and as in discrete time, eq. B-62). Similarly, one can readily show that at the singular strategy, disruptive selection can be characterised from the second-order derivative of  $R_0(z)$ , having

$$h(z^*) = \frac{1}{T^*} \left. \frac{d^2 R_0(z)}{d\epsilon^2} \right|_{\epsilon=0} = \frac{1}{T^*} \left. \frac{\partial^2 R_0(z)}{\partial z_m^2} \right|_{z^*}, \quad (\text{B-92})$$

(see eq. B-63 for the discrete time equivalent of eq. B-92).

Our goal is to reveal how selection depends on changes in state, i.e. unpack the derivatives of fecundity and survival with respect to mutant effect in eqs. (B-84) and (B-88) and obtain eqs. (19), (20) and (II.A) of the main



text. Since selection can be characterised from  $R_0(\mathbf{z})$ , we will work from  $R_0(\mathbf{z})$  following Appendix B.1 of [8] (rather than from eqs. (B-84) and (B-88)). Before doing so in sections B.3.4 and B.3.5, some definitions and preliminaries are in order, which we cover in the next two sections.

### B.3.2 State variables, neutral future reproductive value and its dynamics

**State variables.** First, let

$$\mathbf{y}(a) = (x(a), l(a)) \quad (\text{B-93})$$

collect the physiological state and survival probability, which we collectively call the state variables, of a mutant individual at age  $a$ . Accordingly,  $\mathbf{y}^\circ(a) = (x^\circ(a), l^\circ(a))$  collects the state variables in a resident individual of age  $a$ .

**Future and current reproductive values.** Second, let us define

$$v_f^\circ(a, \mathbf{y}^\circ(a)) = \int_a^M b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt, \quad (\text{B-94})$$

which gives the contribution from age  $a$  onward to individual fitness in a resident, which is closely connected to reproductive number  $R_0$  (eq. B-89). In fact,  $v_f^\circ(a, \mathbf{y}^\circ(a))$  from birth ( $a = 0$ ) reduces to

$$v_f^\circ(0, \mathbf{y}^\circ(0)) = \int_0^M b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt = R_0(\mathbf{z}^\circ) = 1, \quad (\text{B-95})$$

where we used the fact that the resident population is at demographic equilibrium so that each individual produces one offspring on average. We refer to  $v_f^\circ(a, \mathbf{y}^\circ(a))$  as the “future reproductive value” in contrast to the “current reproductive value”, which here is given by

$$\tilde{v}^\circ(a, x^\circ(a)) = \int_a^M \frac{l^\circ(t)}{l^\circ(a)} b(\mathbf{z}^\circ, x^\circ(t)) dt = \frac{v_f^\circ(a, \mathbf{y}^\circ(a))}{l^\circ(a)} \quad (\text{B-96})$$

(eq. 18 in the main text).

**Dynamics of future reproductive values.** To characterise the cascading effects on fitness of a state change at a given age, it will turn out to be useful to quantify the change of future reproductive value  $v_f^\circ(a, \mathbf{y}(a))$  as an individual ages. To that end, consider a small but positive age interval  $\Delta a$ , for which we can write eq. (B-94) as

$$\begin{aligned} v_f^\circ(a, \mathbf{y}^\circ(a)) &= \int_a^{a+\Delta a} b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt + \int_{a+\Delta a}^M b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt \\ &= \int_a^{a+\Delta a} b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt + v_f^\circ(a + \Delta a, \mathbf{y}^\circ(a) + \Delta \mathbf{y}^\circ(a)), \end{aligned} \quad (\text{B-97})$$

where  $\Delta \mathbf{y}^\circ(a) = \mathbf{y}^\circ(a + \Delta a) - \mathbf{y}^\circ(a)$  are changes in state variables over  $\Delta a$ . To the first order in  $\Delta a$ , the first term in the last line of eq. (B-97) reads as

$$\int_a^{a+\Delta a} b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt = b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) \Delta a + \mathcal{O}(\Delta a^2). \quad (\text{B-98})$$

Taking a first-order Taylor expansion of the second term in the last line of eq. (B-97) around  $\Delta a = 0$ , meanwhile, leads to

$$v_f^\circ(a + \Delta a, \mathbf{y}^\circ(a) + \Delta \mathbf{y}^\circ(a)) = v_f^\circ(a, \mathbf{y}^\circ(a)) + \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} \Delta a + \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \Delta \mathbf{y}^\circ(a) + \mathcal{O}(\Delta a^2), \quad (\text{B-99})$$

where

$$\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) = \underbrace{\left( \frac{\partial}{\partial x^\circ(a)}, \frac{\partial}{\partial l^\circ(a)} \right)}_{\equiv \nabla} v_f^\circ(a, \mathbf{y}^\circ(a)) = \left( \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial x^\circ(a)}, \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial l^\circ(a)} \right), \quad (\text{B-100})$$

is the gradient of  $v_f^\circ(a, \mathbf{y}^\circ(a))$  with respect to  $\mathbf{y}^\circ(a)$  ( $\nabla$  is defined as the gradient operator with respect to  $\mathbf{y}^\circ(a)$ ). In optimal control theory,  $\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  are known as **costate variables**, associated respectively with the state variables  $\mathbf{y}(a)$ . Costate variables can be interpreted as giving the marginal effect on residual fitness when (infinitesimally) increasing the associated state variable at age  $a$ .

Plugging eqs. (B-98) and (B-99) into eq. (B-97) yields after some straightforward re-arrangements,

$$-\frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} = b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) + \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \frac{\Delta \mathbf{y}^\circ(a)}{\Delta a} + \mathcal{O}(\Delta a^2). \quad (\text{B-101})$$

Now, using eqs. (16)–(17), we have that

$$\lim_{\Delta a \rightarrow 0} \frac{\Delta \mathbf{y}^\circ(a)}{\Delta a} = \frac{d\mathbf{y}^\circ(a)}{da} = (g(\mathbf{z}^\circ, x^\circ(a)), -\mu(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a)) \equiv \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)). \quad (\text{B-102})$$

Thus, in the limit of  $\Delta a \rightarrow 0$ , eq. (B-101) reads as,

$$\frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} = -\left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) + \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right), \quad (\text{B-103})$$

which is a partial differential equation for  $v_f^\circ(a, \mathbf{y}^\circ(a))$  (this has been derived in the more general context of group-structured populations, eq. B.7 of Appendix of [8] for further discussion). The initial condition of this partial differential equation is given by eq. (B-95), i.e.  $v_f^\circ(0, \mathbf{y}^\circ(0)) = 1$ . The final condition for eq. (B-103) can be obtained by taking the limit  $a \rightarrow M \rightarrow \infty$  of eq. (B-94). Under most realistic biological scenarios (i.e. under external causes of mortality and fecundity limited by environmental constraints), this limit converges to

$$\lim_{a \rightarrow \infty} v_f^\circ(a, \mathbf{y}^\circ(a)) = \lim_{a \rightarrow \infty} \int_a^\infty b(\mathbf{z}^\circ, x^\circ(t)) l^\circ(t) dt = 0. \quad (\text{B-104})$$

Eqs. (B-103), eq. (B-95), and (B-104) say that when a resident individual is born, its future reproductive value is one, which then declines at a rate  $-\left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) + \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right)$  to eventually reach zero.

Solving a partial differential equation is cumbersome, at best. Fortunately, a solution to eq. (B-103) is not always necessary to characterise directional and disruptive selection. In fact, it is only necessary when trait expression depends on state variable(s) (i.e. for plastic traits, see Box I and section B.3.7; [8] for further detail). Otherwise, one only needs to characterise how costate variables  $\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  change with age, which we turn to in the next section.

### B.3.3 Costate variables: dynamics and interpretation

The aim of this section is to derive an ordinary differential equation for costate variables  $\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  (eq. B-100) with respect to age. For this, we first take a total derivative of  $\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  with respect to  $a$  and obtain

$$\frac{d\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))}{da} = \frac{\partial \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))}{\partial a} + \nabla \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \frac{d\mathbf{y}^\circ(a)}{da}. \quad (\text{B-105})$$

It follows from eq. (B-100) that  $\nabla \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  is the Hessian matrix of  $v_f^\circ(a, \mathbf{y}^\circ(a))$ , i.e.

$$\nabla \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) = \nabla \left( \nabla v_f^\circ(a, \mathbf{y}^\circ(a)) \right) = \mathcal{H} \left( v_f^\circ(a, \mathbf{y}^\circ(a)) \right) = \begin{bmatrix} \frac{\partial^2 v_f^\circ(a, x^\circ(a), l^\circ(a))}{\partial x^\circ(a)^2} & \frac{\partial^2 v_f^\circ(a, x^\circ(a), l^\circ(a))}{\partial x^\circ(a) \partial l^\circ(a)} \\ \frac{\partial^2 v_f^\circ(a, x^\circ(a), l^\circ(a))}{\partial l^\circ(a) \partial x^\circ(a)} & \frac{\partial^2 v_f^\circ(a, x^\circ(a), l^\circ(a))}{\partial l^\circ(a)^2} \end{bmatrix}. \quad (\text{B-106})$$

Substituting eq. (B-106) into eq. (B-105), we obtain

$$\frac{d\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))}{da} = \nabla \left( \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} \right) + \mathcal{H} \left( v_f^\circ(a, \mathbf{y}^\circ(a)) \right) \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)), \quad (\text{B-107})$$

where we used eqs. (B-102) and (B-100). To unpack the first term of the right-hand-side of eq. (B-107), let us take the gradient of eq. (B-103) with respect to  $\mathbf{y}^\circ(a)$ , i.e.

$$\nabla \left( \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} \right) = -\nabla \left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) \right) - \nabla \left( \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right). \quad (\text{B-108})$$

Results from vector calculus imply that the gradient of a dot product of two vector-valued functions  $\mathbf{a}$  and  $\mathbf{b}$  can be expressed as  $\nabla(\mathbf{a} \cdot \mathbf{b}) = (\nabla \mathbf{a})^\top \mathbf{b} + (\nabla \mathbf{b})^\top \mathbf{a}$ , where  $\nabla \mathbf{a}$  and  $\nabla \mathbf{b}$  are Jacobian matrices,  $\top$  denotes the transpose operator, and  $(\nabla \mathbf{a})^\top \mathbf{b}$  and  $(\nabla \mathbf{b})^\top \mathbf{a}$  denotes that the transpose of Jacobian matrix  $\nabla \mathbf{a}$  (or  $\nabla \mathbf{b}$ ) is multiplied with the vector  $\mathbf{b}$  (or  $\mathbf{a}$ ). One can easily check this relation holds using Mathematica (e.g. for two-dimensional vector-valued functions like we have here). Using the above relation from vector calculus and eq. (B-106), the second term on the left-hand of eq. (B-108) can be written as

$$\begin{aligned} \nabla \left( \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \cdot \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right) &= \left( \nabla \left( \nabla v_f^\circ(a, \mathbf{y}^\circ(a)) \right) \right)^\top \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) + \left( \nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right)^\top \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \\ &= \left( \mathcal{H} \left( v_f^\circ(a, \mathbf{y}^\circ(a)) \right) \right)^\top \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) + \left( \nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right)^\top \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) \end{aligned} \quad (\text{B-109})$$

where

$$\nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) = \begin{bmatrix} \frac{\partial g(\mathbf{z}^\circ, x^\circ(a))}{\partial x^\circ(a)} & \frac{\partial g(\mathbf{z}^\circ, x^\circ(a))}{\partial l^\circ(a)} \\ -\frac{\partial [\mu(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a)]}{\partial x^\circ(a)} & -\frac{\partial [\mu(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a)]}{\partial l^\circ(a)} \end{bmatrix} \quad (\text{B-110})$$

is the Jacobian matrix of  $\boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a))$ .

Substituting eq. (B-109) into eq. (B-108) yields

$$\nabla \left( \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial a} \right) = -\nabla \left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) \right) - \mathcal{H} \left( v_f^\circ(a, \mathbf{y}^\circ(a)) \right) \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) - \left( \nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a)) \right)^\top \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)), \quad (\text{B-111})$$

where we used the fact that a Hessian matrix is symmetric so that it is equal to its own transpose. Finally, substituting eq. (B-111) into eq. (B-107) and simplifying yields an ordinary differential equation,

$$\frac{d\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))}{da} = -\nabla \left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) \right) - \nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a))^\top \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)). \quad (\text{B-112})$$

For a broad range of biologically feasible conditions ([9] for a formal argument), the final condition to this differential equation is

$$\boldsymbol{\kappa}^\circ(M, \mathbf{y}^\circ(M)) = (0, 0). \quad (\text{B-113})$$

Eqs. (B-112) and (B-113) are necessary in characterising directional and disruptive selection, as shown in sections below B.3.4 and B.3.5.

**Connection with current reproductive value.** Costate variables  $\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  give the marginal effect on future reproductive value  $v_f^\circ(a, \mathbf{y}^\circ(a))$  when increasing the associated state variable  $\mathbf{y}^\circ(a)$  (eq. B-100). These costate variables can be usefully connected to current reproductive value  $\tilde{v}^\circ(a, x^\circ(a))$  via eq. (B-96) ([9] for original connection). In fact, eq. (B-96) reveals that

$$\begin{aligned} \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial x^\circ(a)} &= \frac{\partial \tilde{v}^\circ(a, x^\circ(a))}{\partial x^\circ(a)} l^\circ(a) \equiv \lambda^\circ(a, \mathbf{y}^\circ(a)) \\ \frac{\partial v_f^\circ(a, \mathbf{y}^\circ(a))}{\partial l^\circ(a)} &= \tilde{v}^\circ(a, x^\circ(a)). \end{aligned} \quad (\text{B-114})$$

The first equation means the marginal effect on future reproductive value of a change in physiological state  $x^\circ(a)$  at age  $a$  is equal to the marginal effect of increasing current reproductive value at age  $a$  weighed by survival to that age. The second equation, meanwhile, means the marginal effect on future reproductive value of increasing survival  $l^\circ(a)$  at age  $a$  is equal to the current reproductive value at age  $a$ . From eq. (B-114), costate variables can thus be expressed as,

$$\boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a)) = (\lambda^\circ(a, \mathbf{y}^\circ(a)), \tilde{v}^\circ(a, x^\circ(a))) = \left( \frac{\partial \tilde{v}^\circ(a, x^\circ(a))}{\partial x^\circ(a)} l^\circ(a), \tilde{v}^\circ(a, x^\circ(a)) \right). \quad (\text{B-115})$$

The dynamical system given by eqs. (B-112) and (B-113) can then be written as

$$\frac{d\lambda^\circ(a, \mathbf{y}^\circ(a))}{da} = -\left(\frac{\partial b(\mathbf{z}^\circ, x^\circ(a))}{\partial x^\circ(a)} - \tilde{v}^\circ(a, x^\circ(a))\frac{\partial \mu(\mathbf{z}^\circ, x^\circ(a))}{\partial x^\circ(a)}\right)l^\circ(a) - \lambda^\circ(a, \mathbf{y}^\circ(a))\frac{\partial g(\mathbf{z}^\circ, x^\circ(a))}{\partial x(a)} \quad (\text{B-116})$$

and

$$\frac{d\tilde{v}^\circ(a, x^\circ(a))}{da} = -\left(b(\mathbf{z}^\circ, x^\circ(a)) - \tilde{v}^\circ(a, x^\circ(a))\mu(\mathbf{z}^\circ, x^\circ(a))\right) \quad (\text{B-117})$$

together with conditions,

$$\begin{aligned} \lambda^\circ(M, \mathbf{y}^\circ(M)) &= 0 \\ \tilde{v}^\circ(M, x^\circ(M)) &= 0. \end{aligned} \quad (\text{B-118})$$

### B.3.4 Directional selection

We are finally in a position to derive eq. (19) of the main text using standard techniques from optimal control theory. Formally, our model here corresponds to a special case of constant controls of optimal control theory, because the evolving trait  $z$  (or the “control variable” in the language of optimal control theory) is a scalar (see e.g. [10, 11] for optimal control textbooks and e.g. [8, 9] for application in evolutionary biology). This section follows closely [9], where the same result has been shown in the more general case of group-structured population.

The crux of the argument is to first add a zero quantity to the basic reproductive number (eq. B-89) as follows

$$R_0(\mathbf{z}) = \int_0^M b(\mathbf{z}, x(a))l(a) da + \underbrace{\int_0^M \boldsymbol{\kappa}^\circ(a) \cdot \left(\boldsymbol{\theta}(\mathbf{z}, \mathbf{y}(a)) - \frac{d\mathbf{y}(a)}{da}\right) da}_{=0} \quad (\text{B-119})$$

using eq. (B-102) and writing  $\boldsymbol{\kappa}^\circ(a) = \boldsymbol{\kappa}^\circ(a, \mathbf{y}^\circ(a))$  for short. The last term of eq. (B-119) can be rewritten using integration by parts as,

$$\begin{aligned} -\int_0^M \boldsymbol{\kappa}^\circ(a) \cdot \frac{d\mathbf{y}(a)}{da} da &= \int_0^M \mathbf{y}(a) \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da} da - \left(\boldsymbol{\kappa}^\circ(M) \cdot \mathbf{y}(M)\right) + \boldsymbol{\kappa}^\circ(0) \cdot \mathbf{y}(0) \\ &= \int_0^M \mathbf{y}(a) \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da} da - \boldsymbol{\kappa}^\circ(0) \cdot \mathbf{y}(0), \end{aligned} \quad (\text{B-120})$$

where  $M \rightarrow \infty$  and we used eq. (B-113). Substituting eq. (B-120) into (B-119) yields

$$R_0(\mathbf{z}) = \int_0^M \left(b(\mathbf{z}, x(a))l(a) + \boldsymbol{\kappa}^\circ(a) \cdot \boldsymbol{\theta}(\mathbf{z}, \mathbf{y}(a)) + \mathbf{y}(a) \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da}\right) da + \boldsymbol{\kappa}^\circ(0) \cdot \mathbf{y}(0). \quad (\text{B-121})$$

We then take the derivative of eq. (B-121) with respect to  $z_m$  and after some re-arrangements obtain,

$$\begin{aligned} \left. \frac{\partial R_0(\mathbf{z})}{\partial z_m} \right|_z &= \int_0^M \left[ \frac{\partial b(\mathbf{z}, x^\circ(a))}{\partial z_m} l^\circ(a) + \boldsymbol{\kappa}^\circ(a) \cdot \frac{\partial \boldsymbol{\theta}(\mathbf{z}, \mathbf{y}^\circ)}{\partial z_m} \right. \\ &\left. + \left\{ \nabla \left( b(\mathbf{z}^\circ, x^\circ(a)) l^\circ(a) \right) + \nabla \boldsymbol{\theta}(\mathbf{z}^\circ, \mathbf{y}^\circ(a))^\top \boldsymbol{\kappa}^\circ(a) + \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right\} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} \right] da. \end{aligned} \quad (\text{B-122})$$

Note however from eq. (B-112) that the term between curly brackets in eq. (B-122) is in fact a zero vector,  $(0, 0)$ .

We can thus write eq. (B-122) as

$$\left. \frac{\partial R_0(\mathbf{z})}{\partial z_m} \right|_z = \int_0^M \left[ \left( \frac{\partial b(\mathbf{z}, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(\mathbf{z}, x^\circ(a))}{\partial z_m} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial g(\mathbf{z}, x^\circ(a))}{\partial z_m} \right] da, \quad (\text{B-123})$$

where we used  $\boldsymbol{\kappa}^\circ(a) = (\lambda^\circ(a), \tilde{v}^\circ(a))$  (eq. B-115 with shortened notation) and  $\boldsymbol{\theta}(\mathbf{z}, \mathbf{y}(a)) = (g(\mathbf{z}, x(a)), -\mu(\mathbf{z}, x(a))l(a))$  (eq. B-102). Substituting eq. (B-123) into eq. (B-91) then yields the selection gradient,

$$s(z) = \frac{1}{T^\circ} \int_0^M \left[ \left( \frac{\partial b(\mathbf{z}, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(\mathbf{z}, x^\circ(a))}{\partial z_m} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial g(\mathbf{z}, x^\circ(a))}{\partial z_m} \right] da. \quad (\text{B-124})$$

Rewriting eq. (B-115) as  $\lambda^\circ(a) = [\partial \tilde{v}^\circ(a) / \partial x^\circ(a)] l^\circ(a)$  (i.e. removing dependence on  $x^\circ(a)$  in the notation) and expanding definitions of  $\mathbf{z} = (z_m, z)$  and  $\mathbf{z}^\circ = (z, z)$ , allow us to express eq. (B-124) as

$$s(z) = \frac{1}{T^\circ} \int_0^M \left( \frac{\partial b(z_m, z, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(z_m, z, x^\circ(a))}{\partial z_m} + \frac{\partial \tilde{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial g(z_m, z, x^\circ(a))}{\partial z_m} \right) l^\circ(a) da, \quad (\text{B-125})$$

which is eq. (19) of the main text. In contrast to eq. (B-84), the partial derivatives in eq. (B-125) ( $\partial b(z_m, z, x^\circ(a)) / \partial z_m$ ,  $\partial \mu(z_m, z, x^\circ(a)) / \partial z_m$ , and  $\partial g(z_m, z, x^\circ(a)) / \partial z_m$ ) are now derivatives with respect to the first argument only (as the state is now evaluated at the resident, i.e. as  $x^\circ(a)$ ) and can thus be straightforwardly computed. Together with the dynamic eqs. (16)–(17) for state variables  $x^\circ(a)$  and  $l^\circ(a)$  and eqs. (B-116)–(B-118) for costate variables  $\tilde{v}^\circ(a)$  and  $\lambda^\circ(a)$ , respectively, eq. (B-125) allows one to characterise directional selection  $s(z)$  and compute singular strategies  $z^*$  such that  $s(z^*) = 0$  (see e.g. [8, 9, 12, 13] for such a procedure).

### B.3.5 Disruptive selection

To derive eqs. (20) and (II.A) of the main text for disruptive selection, we follow a similar argument as the one used above for directional selection. To that end, it is convenient to use the Hamiltonian function,

$$H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a)) = \left( b(\mathbf{z}, x(a)) - \tilde{v}^\circ(a) \mu(\mathbf{z}, x(a)) \right) l(a) + \lambda^\circ(a) g(\mathbf{z}, x(a)), \quad (\text{B-126})$$

which can be thought of as giving the increase in fitness at age  $a$  from different activities (e.g. reproducing, surviving and changing the internal state). Using the Hamiltonian, we can rewrite eq. (B-121) as

$$R_0(\mathbf{z}) = \int_0^M \left( H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a)) + \mathbf{y}(a) \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right) da + \boldsymbol{\kappa}^\circ(0) \cdot \mathbf{y}(0). \quad (\text{B-127})$$

Taking a second-order derivative of eq. (B-127) with respect to  $z_m$  yields

$$\begin{aligned} \left. \frac{\partial^2 R_0(\mathbf{z})}{\partial z_m^2} \right|_z = \int_0^M \left[ \frac{\partial^2 H(\mathbf{z}, \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m^2} + \left( \frac{\partial H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)} + \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right) \cdot \frac{\partial^2 \mathbf{y}(a)}{\partial z_m^2} \right. \\ \left. + 2 \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial \mathbf{y}(a)} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} + \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)^2} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} \right] da \end{aligned} \quad (\text{B-128})$$

Note however that from eq. (B-126) (and eqs. B-116 and B-117), we have

$$\left. \frac{\partial H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)} \right|_z = - \left. \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right|_z. \quad (\text{B-129})$$

The term within brackets on the first line of eq. (B-128) thus vanishes, so that eq. (B-128) simplifies to,

$$\begin{aligned} \left. \frac{\partial^2 R_0(\mathbf{z})}{\partial z_m^2} \right|_z = \int_0^M \left[ \frac{\partial^2 H(\mathbf{z}, \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m^2} \right. \\ \left. + 2 \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial \mathbf{y}(a)} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} + \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)^2} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m} \right] da. \end{aligned} \quad (\text{B-130})$$

Next, we substitute for the matrix

$$\begin{aligned} \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)^2} &= \begin{bmatrix} \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a)^2} & \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a) \partial l(a)} \\ \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial l(a) \partial x(a)} & \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial l(a)^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a)^2} & \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a) \partial l(a)} \\ \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial l(a) \partial x(a)} & 0 \end{bmatrix} \end{aligned} \quad (\text{B-131})$$

(where we used the definition of the Hamiltonian eq. B-126 to obtain  $\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a)) / (\partial l(a)^2) = 0$ ), and the vectors  $\mathbf{y} = (x(a), l(a))$  and  $\boldsymbol{\kappa}^\circ(a) = (\lambda^\circ(a), \bar{v}^\circ(a))$  into eq. (B-130), to get

$$\begin{aligned} \left. \frac{\partial^2 R_0(\mathbf{z})}{\partial z_m^2} \right|_{z^*} = \int_0^M \left[ \frac{\partial^2 H(\mathbf{z}, \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m^2} + 2 \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial l(a)} \frac{\partial l(a)}{\partial z_m} + 2 \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a) \partial l(a)} \frac{\partial x(a)}{\partial z_m} \frac{\partial l(a)}{\partial z_m} \right. \\ \left. + 2 \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial x(a)} \frac{\partial x(a)}{\partial z_m} + \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a)^2} \left( \frac{\partial x(a)}{\partial z_m} \right)^2 \right] da. \end{aligned} \quad (\text{B-132})$$

Now, from the definition of the Hamiltonian eq. (B-126), the various second-order derivatives appearing in eq. (B-132) are,

$$\begin{aligned}
\left. \frac{\partial^2 H(\mathbf{z}, \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m^2} \right|_z &= \left( \frac{\partial^2 b(\mathbf{z}, x^\circ(a))}{\partial z_m^2} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(\mathbf{z}, x^\circ(a))}{\partial z_m^2} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial^2 g(\mathbf{z}, x^\circ(a))}{\partial z_m^2} \Big|_z, \\
\left. \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial l(a)} \right|_z &= \left( \frac{\partial b(\mathbf{z}, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(\mathbf{z}, x^\circ(a))}{\partial z_m} \right) \Big|_z \\
\left. \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a) \partial l(a)} \right|_z &= \left( \frac{\partial b(\mathbf{z}^\circ, x(a))}{\partial x(a)} - \tilde{v}^\circ(a) \frac{\partial \mu(\mathbf{z}^\circ, x(a))}{\partial x(a)} \right) \Big|_z \tag{B-133} \\
\left. \frac{\partial^2 H(\mathbf{z}, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m \partial x(a)} \right|_z &= \left( \frac{\partial^2 b(\mathbf{z}, x(a))}{\partial z_m \partial x(a)} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(\mathbf{z}, x(a))}{\partial z_m \partial x(a)} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial^2 g(\mathbf{z}, x(a))}{\partial z_m \partial x(a)} \Big|_z \\
\left. \frac{\partial^2 H(\mathbf{z}^\circ, \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial x(a)^2} \right|_z &= \left( \frac{\partial^2 b(\mathbf{z}^\circ, x(a))}{\partial x(a)^2} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(\mathbf{z}^\circ, x(a))}{\partial x(a)^2} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial^2 g(\mathbf{z}^\circ, x(a))}{\partial x(a)^2} \Big|_z.
\end{aligned}$$

Substituting eq. (B-133) into eq. (B-132), which is turn substituted into eq. (B-92), we find that using notation from the main text, disruptive selection can be expressed as

$$h(z^*) = h_w(z^*) + 2h_q(z^*) \tag{B-134}$$

where

$$\begin{aligned}
h_w(z^*) &= \frac{1}{T^*} \int_0^M \left[ \left( \frac{\partial^2 b(z_m, z, x^\circ(a))}{\partial z_m^2} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(z_m, z, x^\circ(a))}{\partial z_m^2} + \frac{\partial^2 g(z_m, z, x^\circ(a))}{\partial z_m^2} \frac{\partial \tilde{v}^\circ(a)}{\partial x^\circ(a)} \right) l^\circ(a) \right]_{z^*} da \\
h_q(z^*) &= \frac{1}{T^*} \int_0^M \left[ \left( \frac{\partial b(z_m, z, x^\circ(a))}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(z_m, z, x^\circ(a))}{\partial z_m} \right) \frac{\partial l(a)}{\partial z_m} + h_{q,x}(a) \frac{\partial x(a)}{\partial z_m} \right]_{z^*} da,
\end{aligned} \tag{B-135}$$

which gives eq. (20) of the main text and

$$\begin{aligned}
h_{q,x}(a) &= \left[ \left( \frac{\partial b(z, z, x(a))}{\partial x(a)} - \tilde{v}^\circ(a) \frac{\partial \mu(z, z, x(a))}{\partial x(a)} \right) \frac{\partial l(a)}{\partial z_m} \right. \\
&\quad + \left( \frac{\partial^2 b(z_m, z, x(a))}{\partial z_m \partial x(a)} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(z_m, z, x(a))}{\partial z_m \partial x(a)} + \frac{\partial \tilde{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial^2 g(z_m, z, x(a))}{\partial z_m \partial x(a)} \right) l^\circ(a) \\
&\quad \left. + \frac{1}{2} \left( \frac{\partial^2 b(z, z, x(a))}{\partial x(a)^2} - \tilde{v}^\circ(a) \frac{\partial^2 \mu(z, z, x(a))}{\partial x(a)^2} + \frac{\partial \tilde{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial^2 g(z, z, x(a))}{\partial x^2} \right) \frac{\partial x(a)}{\partial z_m} l^\circ(a) \right]_{z^*},
\end{aligned} \tag{B-136}$$

which gives eq. (II.A) of Box II of the main text.

### B.3.6 Directional and disruptive selection under continuous age-structure

Eqs. (B-125) and (B-135) can readily reduce to the case where age is continuous (i.e. to the continuous versions of eqs. 14 and 15 of the main text). If physiological state  $x(a) = x_c$  is constant, then  $g(\mathbf{z}, x) = 0$  and the fecundity rate and mortality rate are just functions of age, i.e.  $b(\mathbf{z}, x_c, a) = b(\mathbf{z}, a)$  and  $\mu(\mathbf{z}, x_c, a) = \mu(\mathbf{z}, a)$ . Substituting these fitness components into eq. (B-124) gives

$$s(z) = \frac{1}{T^\circ} \int_0^M \left( \frac{\partial b(\mathbf{z}, a)}{\partial z_m} - \tilde{v}^\circ(a) \frac{\partial \mu(\mathbf{z}, a)}{\partial z_m} \right) l^\circ(a) da, \tag{B-137}$$



which is the continuous time version of eq. (14). Similarly, substituting these fitness components into eq. (B-135) leads to

$$\begin{aligned} h_w(z^*) &= \frac{1}{T^*} \int_0^M \left[ \left( \frac{\partial^2 b(z, a)}{\partial z_m^2} - \bar{v}^\circ(a) \frac{\partial^2 \mu(z, a)}{\partial z_m^2} \right) l^\circ(a) \right]_{z^*} da, \\ h_q(z^*) &= \frac{1}{T^*} \int_0^M \left[ \left( \frac{\partial b(z, a)}{\partial z_m} - \bar{v}^\circ(a) \frac{\partial \mu(z, a)}{\partial z_m} \right) \frac{\partial l(a)}{\partial z_m} \right]_{z^*} da, \end{aligned} \quad (\text{B-138})$$

which are continuous time versions of eq. 15.

### B.3.7 Directional selection of phenotypically plastic traits

In this section, we derive the expression for directional selection gradient for phenotypically plastic traits, shown in Box I of the main text. Here, phenotypically plastic traits mean traits whose expression can be age- and state-dependent. We discuss the the dynamic properties of the costate variable and reproductive value under different assumptions about trait expression. Our aim here is to highlight some crucial steps in the derivations and outline the main differences in the properties of directional selection under various conceptualisations of traits. For more details about age- and state-dependent traits and their relationship under more broader context of group-structured populations (which subsumes well-mixed population as a special case), see [8].

In this context, it is useful to distinguish between (i) trait expressions (or actions); namely, measurement of a characteristics of an organism throughout the lifespan of individuals (e.g. allocation to growth at different ages) and (ii) the traits; namely, functional characteristics of organisms that can be thought of as decision rules or strategies (e.g. allocation to growth as a function of nutrition level). Henceforward, we refer to the first term as the *trait expression* and the second terms as the *control*. We first discuss trait expressions and then specify the relationship between trait expressions and controls. We denote by  $z_m = \{z_m(a)\}_{a \in [0, M]}$  and  $z = \{z(a)\}_{a \in [0, M]}$ , the trait expressions over their entire life course (i.e. schedules) for mutant and resident, respectively. The mutant trait expression can be written as the deviation from the resident at any age  $a$  as

$$z_m(a) = z(a) + \epsilon \xi(a), \quad (\text{B-139})$$

where  $0 > \epsilon \ll 1$  is a small parameter tuning the effect size of the deviation and  $\xi(a)$  is the admissible trait deviation at age  $a$ , such that  $z_m(a) \in Z = [z_{\min}, z_{\max}] \forall a \in [0, M]$ , where  $z_{\min}$  and  $z_{\max} \in \mathbb{R}$  give the minimum value and maximum values for the trait expression. This means that for a candidate uninadable strategy  $z^* = \{z^*(a)\}_{a \in [0, M]}$  (recall eq. 1) the admissible trait deviations are given for each age  $a$  as follows

$$\begin{aligned} z^*(a) = z_{\min} & & \xi(a) \geq 0, \\ z_{\min}(a) < z^*(a) < z_{\max}(a) & & \xi(a) \text{ unrestricted}, \\ z^*(a) = z_{\max}(a) & & \xi(a) \leq 0. \end{aligned} \quad (\text{B-140})$$

If the trait space is unbounded (i.e.  $Z = \mathbb{R}$ ), then the second line gives the admissible trait deviations for all ages  $a$ . We emphasise the properties of bounded controls, because the bounds become especially relevant for age- and state-dependent traits (e.g. bang-bang strategies of growth and reproduction, where the trait  $z(a) \in [0, 1]$  is the proportional resource allocation to growth versus reproduction).

There are at least three useful conceptualisations of controls that are relevant to evolutionary biology, which can be expressed as

$$z(a) = \begin{cases} u(a, x^\circ(a)) \text{ closed-loop or feedback control (i.e. state-dependent trait),} \\ u(a) \text{ open-loop control (i.e. age-dependent trait),} \\ u_c \text{ constant control (i.e. fixed trait).} \end{cases} \quad (\text{B-141})$$

Thus far, our conceptualisation of trait expression formally corresponds to the case of constant controls. In life-history theory, traits are often conceptualised either as open-loop controls (i.e. age-dependent traits, e.g. age-specific allocation strategies to growth vs reproduction) or piece-wise constant controls (e.g. age-at-maturity models, where bang-bang schedules are assumed). In behavioural ecology we often see closed-loop controls (i.e. internal state-dependent traits, e.g. survival of winter model in foraging theory). Closed-loop control is the most general of these conceptualisations that subsumes the other conceptualisations. We now discuss selection on traits conceptualised as closed-loop controls.

Let us first note that  $R_0(\mathbf{z})$  under closed-loop formalisation can be expressed as

$$R_0(\mathbf{z}) = \int_0^\infty b(a, \mathbf{z}(a), x(a)) l(a) da, \quad (\text{B-142})$$

where the fecundity rate  $b(a, \mathbf{z}(a), x(a))$  depends on trait expression  $\mathbf{z}(a) = \mathbf{u}(a, \mathbf{x}(a)) = (u(a, x(a)), u(a, x^\circ(a)))$  at age  $a$ , where  $\mathbf{x}(a) = (x(a), x^\circ(a))$  and here the fecundity rate can also depend directly on age  $a$  (e.g. allowing for age-dependent availability of resources). The dynamic constraints imposed by the state variables can be similarly express as follows

$$\frac{dx(a)}{da} = g(a, \mathbf{z}(a), x(a)) \quad \text{with } x(0) = x_b, \quad \frac{dl(a)}{da} = -\mu(a, \mathbf{z}(a), x(a)) l(a), \quad \text{with } l(0) = 1. \quad (\text{B-143})$$

Let us now adjoin the basic reproductive number as we did in sections B.3.4–B.3.5 (especially, recall eq. B-121 and eq. B-126), which allows us to end up with an adjoined basic reproductive number in the following form

$$R_0(\mathbf{z}) = \int_0^M \left( H(a, \mathbf{z}(a), \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a)) + \mathbf{y}(a) \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right) da + \boldsymbol{\kappa}^\circ(0) \cdot \mathbf{y}(0) \quad (\text{B-144})$$

where we used the definition of the Hamiltonian function (recall eq. B-126 and that  $\mathbf{y}(a) = (x(a), l(a))$  and

$\boldsymbol{\kappa}^\circ(a) = (\lambda^\circ(a), \bar{v}^\circ(a))$ ). Note that the Hamiltonian under closed-loop controls takes the form

$$\begin{aligned} H(a) \equiv H(a, \mathbf{z}(a), \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a)) &= H(a, \mathbf{z}(a), x(a), l(a), \lambda^\circ(a), \bar{v}^\circ(a)) = \\ & \left( b(a, \mathbf{z}(a), x(a)) - \mu(a, \mathbf{z}(a), x(a)) \bar{v}^\circ(a) \right) l(a) + \lambda^\circ(a) g(a, \mathbf{z}(a), x(a)). \end{aligned} \quad (\text{B-145})$$

Here the selection gradient (recall eq. B-91) takes a form of a so-called *Gâteaux derivative* (Section 3 in [14]) of the basic reproductive number (given by eq. B-144),

$$s(z) = \frac{1}{T^\circ} \left. \frac{dR_0(\mathbf{z})}{d\epsilon} \right|_{\epsilon=0} = \int_0^M \left. \frac{\partial R_0(\mathbf{z})}{\partial z_m(a')} \right|_z \cdot \xi(a') da', \quad (\text{B-146})$$

where the “ $\cdot$ ” here denotes the inner product of two functions (see e.g. [15] Chapter 6) and here (and throughout this section) evaluation at  $z$  means that everything has been evaluated at the (closed-loop) resident control  $z_m = z = u(x^\circ)$ , where  $u(x^\circ) = \{u(a, x^\circ(a))\}_{a \in [0, M]}$ . Using eq. (B-144) and the definition and properties of the Gâteaux derivatives (eqs. A.1 and A.2. in Appendix A.1 in [8]) which can be expressed in terms of point-wise variations, which yields

$$\begin{aligned} s(z) &= \int_0^M \int_0^M \left[ \frac{\partial H(a, \mathbf{z}(a), \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m(a')} + \frac{\partial \mathbf{y}(a)}{\partial z_m(a')} \cdot \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right]_z \cdot \xi(a') da da' \\ &= \int_0^M \int_0^M \left[ \frac{\partial H(a, \mathbf{z}(a), \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m(a)} \frac{\partial z_m(a)}{\partial z_m(a')} \right. \\ & \quad \left. + \underbrace{\left( \frac{\partial H(a, \mathbf{z}^\circ(a), \mathbf{y}(a), \boldsymbol{\kappa}^\circ(a))}{\partial \mathbf{y}(a)} + \frac{d\boldsymbol{\kappa}^\circ(a)}{da} \right)}_{(0,0)} \cdot \frac{\partial \mathbf{y}(a)}{\partial z_m(a')} \right]_z \cdot \xi(a') da da' \\ &= \int_0^M \int_0^M \left. \frac{\partial H(a, \mathbf{z}(a), \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m(a)} \frac{\partial z_m(a)}{\partial z_m(a')} \right|_z \cdot \xi(a') da da' \\ &= \int_0^M \int_0^M \left. \frac{\partial H(a, \mathbf{z}(a), \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m(a)} \delta_a(a') \right|_z \cdot \xi(a) da da', \\ &= \int_0^M \left. \frac{\partial H(a, \mathbf{z}(a), \mathbf{y}^\circ(a), \boldsymbol{\kappa}^\circ(a))}{\partial z_m(a)} \right|_z \cdot \xi(a) da, \\ &= \frac{1}{T^\circ} \int_0^M \left[ \left( \frac{\partial b(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} - \bar{v}^\circ(a) \frac{\partial \mu(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right) l^\circ(a) + \lambda^\circ(a) \frac{\partial g(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right]_z \cdot \xi(a) da \end{aligned} \quad (\text{B-147})$$

For the first equality we used eq. (B-144) in eq. (B-146). For the second equality we used the chain and product rules. For the third equality we used that  $-d\boldsymbol{\kappa}^\circ(a)/da|_z = \partial H(a)/(\partial \mathbf{y}(a))|_z$  (recall eqs. B-116 and B-117 and eq. B-145). For the fourth equality, we used the fact that trait deviations at different ages are independent from each other (i.e.  $\partial z_m(a)/(\partial z_m(a')) = 0$  for all  $a' \neq a$ ) and here  $\delta_a(a')$  denotes the Dirac delta function peaked at  $a$  with  $\delta_a(a')$  being the value of the function at  $a'$  ( $\delta_a(a') = 0$  for all  $a' \neq a$ ). The exterior integral thus vanishes whenever  $a' \neq a$  (fifth equality). Finally, for the sixth equality we substituted the Hamiltonian (B-145). Recall from the first line of eq. (B-115) that  $\lambda^\circ(a) = \lambda^\circ(a, x^\circ(a), l^\circ(a)) = \partial \bar{v}^\circ(a, x^\circ(a), l^\circ(a))/(\partial x^\circ(a)) l^\circ(a)$ , which upon substitution into eq. (B-147) yields that the selection gradient under closed-loop and open-loop trait

expression takes the following form

$$s(z) = \frac{1}{T^\circ} \int_0^M \left( \frac{\partial b(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} - \bar{v}^\circ(a) \frac{\partial \mu(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} + \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial g(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right) l^\circ(a) \cdot \xi(a) da \quad (\text{B-148})$$

(see also eqs. 26 and 37 together with eq. 6 of [8]). We now outline the main difference between the selection gradient (B-148) for closed-loop and open-loop traits and selection gradient (19) and discuss the difference between closed-loop and open-loop conceptualisations.

First, we note that the directional selection gradient  $s(z)$  can be decomposed into age-specific selection gradient  $s(z, a) = \partial H(a) / (\partial z_m(a))$  (i.e.  $s(z) = \int_0^M s(z, a) \cdot \xi(a) da$ ), which specifies the direction of selection for each age  $a$ . In particular, eq. (B-140) yields that the selection gradient  $s(z^*, a)$  for an uninvadable strategy  $z^*$  must necessarily satisfy

$$\begin{aligned} z^*(a) &= z_{\min} && \text{only if } s(z^*, a) \leq 0 \\ z_{\min}(a) &< z^*(a) < z_{\max}(a) && \text{only if } s(z^*, a) = 0 \\ z^*(a) &= z_{\max}(a) && \text{only if } s(z^*, a) \geq 0. \end{aligned} \quad (\text{B-149})$$

This means that for singular traits ( $s(z^*, a) = 0$ ) an age-specific balance condition must hold

$$\left. \frac{\partial b(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right|_{z^*} = \bar{v}^\circ(a) \left. \frac{\partial \mu(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right|_{z^*} - \left. \frac{\partial \bar{v}^\circ(a)}{\partial x^\circ(a)} \frac{\partial g(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right|_{z^*}. \quad (\text{B-150})$$

Intuitively, eq. (B-150) says that increasing fecundity at age  $a$  comes at the cost of increase in mortality weighed by its effect on future fitness and decreases in the condition of internal state weighed by its effect on future fitness. Importantly, the trade-offs between fecundity, survival and internal state is independent on survival of the individual to the current age. This feature comes from the fact that trait expression can change at different ages, and thus this feature arises under both closed-loop and open-loop control conceptualisations.

Second, the dynamics of the costate variable  $\lambda^\circ(a) = \partial \bar{v}^\circ(a) / (\partial x^\circ(a)) \times l^\circ(a) = \partial \bar{v}^\circ(a, x^\circ(a), l^\circ(a)) / (\partial x^\circ(a)) \times l^\circ(a)$  (recall eq. B-115) contains an additional term that accounts for the feedback of state change on control. This can be seen from eq. (B-116) (recalling that we have short-hand notations  $\lambda^\circ(a) = \lambda^\circ(a, x^\circ(a), l^\circ(a))$  and  $\bar{v}^\circ(a) = \bar{v}^\circ(a, x^\circ(a))$ ) and noting that here  $b(\mathbf{z}, x(a)) = b(a, \mathbf{z}(a), x(a))$ ,  $\mu(\mathbf{z}, x(a)) = \mu(a, \mathbf{z}(a), x(a))$ , and  $\theta(\mathbf{z}, x(a)) = \theta(a, \mathbf{z}(a), x(a))$ , then it follows that the dynamics of the costate variable  $\lambda^\circ(s)$  is given by

$$\begin{aligned} -\frac{d\lambda^\circ(a)}{da} &= \left( \frac{\partial b(a, \mathbf{u}(a, x(a)), x(a))}{\partial x(a)} - \bar{v}^\circ(a) \frac{\partial \mu(a, \mathbf{u}(a, x(a)), x(a))}{\partial x(a)} \right) l^\circ(a) - \lambda^\circ(a) \frac{\partial g(a, \mathbf{u}(a, x(a)), x(a))}{\partial x(a)} \Big|_z, \\ &= \left[ \left( \frac{\partial b(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} - \bar{v}^\circ(a) \frac{\partial \mu(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right) l^\circ(a) - \lambda^\circ(a) \frac{\partial g(a, \mathbf{z}(a), x^\circ(a))}{\partial z_m(a)} \right] \frac{\partial u(a, x(a))}{\partial x(a)} \Big|_z, \\ &+ \left( \frac{\partial b(a, \mathbf{z}^\circ(a), x(a))}{\partial x(a)} - \bar{v}^\circ(a) \frac{\partial \mu(a, \mathbf{z}^\circ(a), x(a))}{\partial x(a)} \right) l^\circ(a) - \lambda^\circ(a) \frac{\partial g(a, \mathbf{z}^\circ(a), x(a))}{\partial x(a)} \Big|_z. \end{aligned} \quad (\text{B-151})$$

where  $\partial u(a, x(a)) / \partial x(a)$  can be interpreted as the trait sensitivity of an individual with respect to changes in its internal state variable  $x(a)$  (see eq. 34 in [8] and therein for a more detailed proof). Using the eq. (B-145) in

eq. (B-151) and re-arranging we can represent the costate dynamics as

$$-\frac{d\lambda^\circ(a)}{da} = \frac{\partial H(a, \mathbf{z}^\circ(a), x(a), l^\circ(a), \lambda^\circ(a), \tilde{v}^\circ(a))}{\partial x(a)} \Big|_z + \underbrace{\frac{\partial H(a, \mathbf{z}(a), x^\circ(a), l^\circ(a), \lambda^\circ(a), \tilde{v}^\circ(a))}{\partial z_m(a)} \Big|_z}_{\text{feedback term}} \frac{\partial u(a, x(a))}{\partial x(a)} \Big|_z. \quad (\text{B-152})$$

The second term in eq. (B-152) can be interpreted as a state-feedback and it appears only under closed-loop conceptualisation and is a fundamental feature that distinguishes closed-loop conceptualisation of trait expression from open-loop and constant control conceptualisations of trait expression. Intuitively, the properties of the feedback term determines the relationship between models with age-dependent trait expression (e.g. age-dependent resource allocation) and state-dependents models (e.g. size-dependent resource allocation), and delineates the conditions under which they lead to different predictions about directional selection.

Due to these feedback effects in the dynamics of the costate variable  $\lambda^\circ(a)$ , eq. (B-152) can not be directly solved, since it implicitly contains higher order derivatives of neutral future reproductive value  $v_f^\circ(a) = v_f^\circ(a, x^\circ(a), l^\circ(a))$  via  $\partial u(a, x(a))/\partial x(a)$  (see also eq. 33 and discussion in [8]). Under closed-loop conceptualisation of traits, the co-state variable  $\lambda^\circ(a)$  can be found by first solving a partial differential equation for future reproductive value (recall eq. B-103), which we can express here as

$$-\frac{\partial v_f^\circ(a, x^\circ(a), l^\circ(a))}{\partial a} = b(a, \mathbf{z}^\circ(a), x^\circ(a))l^\circ(a) + \nabla v_f^\circ(a, x^\circ(a), l^\circ(a)) \cdot \boldsymbol{\theta}(a, \mathbf{z}^\circ(a), \mathbf{y}^\circ(a)) \quad (\text{B-153})$$

and then taking the derivative to find  $\lambda^\circ(a)$  as  $\lambda^\circ(a) = \lambda^\circ(a, x^\circ(a), l^\circ(a)) = \partial v_f^\circ(a, x^\circ(a), l^\circ(a))/(\partial x)|_z$ .

Finally, we would like to note that for an internal singular strategy  $z^*(a)$  for which  $s(z^*, a) = \partial H(a)/z_m(a)|_{z^*} = 0$ , the multiplier of  $\partial u(a, x(a))/\partial x(a)$  is zero. Hence, the feedback effect in co-state dynamics  $\lambda^\circ(a)$  vanishes and thus closed-loop and open-loop conceptualisations of controls yield the same evolutionary outcome. Loosely speaking, the feedback effects are non-zero when there are genetic conflicts between individuals in interaction. Thus closed-loop conceptualisation is relevant in locally subdivided populations, which has been worked out in [8].

## C Selection in subdivided populations

### C.1 Homogeneous groups

Here we consider the case where individuals within groups are homogeneous and derives eqs. (21)-(23) of the main text.

### C.1.1 Life-cycle events

We follow [16] and consider a haploid population divided into an infinite number of groups each with  $n$  adult individuals with the following life-cycle events: (i) groups may go extinct, in which case all individuals present in the group die; (ii) each of the  $n$  adults in surviving groups produces offspring (in sufficient numbers for each group to always be of size  $n$  at the beginning of stage 1 of the life cycle) and then either survives or dies; finally, (iii) dispersal and density-dependent competition for vacated breeding spots occur. This life cycle allows for one, several, or all adults to die per life cycle iteration so allowing for overlapping and non-overlapping generations as well as metapopulation processes where whole groups go extinct and get re-colonized. We assume that each offspring has a nonzero probability of dispersal so that groups are not isolated from one another. Dispersal may occur in groups (propagule dispersal) and before or after density-dependent competition.

### C.1.2 Fitness

In a population subdivided among homogeneous groups of size  $n$ , the dynamics of the mutant are modelled by considering the  $n \times 1$  vector  $\mathbf{N}_t = (N_{1,t}, N_{2,t}, \dots, N_{n,t})$ , consisting of the variables  $N_{k,t}$  that give the number of groups with  $k$  mutants at some time  $t$  when the mutant is rare. The dynamics of this vector are given by,

$$\mathbf{N}_{t+1} = \mathbf{A}(z_m, z) \cdot \mathbf{N}_t, \quad (\text{C-1})$$

where the  $n \times n$  matrix  $\mathbf{A}(z_m, z)$  has  $(i, j)$  entry  $a_{ij}(z_m, z)$  equal to the expected number of groups with  $1 \leq i \leq n$  mutants produced by a focal group with  $1 \leq j \leq n$  mutants (including the focal group if it transits between number of mutants). Such a notion of reproduction at the group level is possible because we assume that the mutant is rare, that the number of groups is large, and that dispersal among groups is random. These assumptions entail that a mutant group (i.e. a group with at least one mutant) cannot receive mutant immigrants, and that a resident group (i.e. a group with no mutant) cannot receive mutant immigrants from different mutant groups. This in turn means that we can clearly ascribe parentage between groups (where the unique parent group is the one who sends a successful immigrant into a resident group).

Invasion fitness is given by the leading eigenvalue  $\rho(z_m, z)$  of the mean matrix  $\mathbf{A}(z_m, z)$ . This eigenvalue satisfies

$$\rho(z_m, z) \mathbf{u}(z_m, z) = \mathbf{A}(z_m, z) \cdot \mathbf{u}(z_m, z) \quad (\text{C-2})$$

where  $\mathbf{u}$  is the right eigenvector of  $\mathbf{A}(z_m, z)$ , scaled such that its entries sum to one (i.e.  $\sum_{i=1}^n u_i(z_m, z) = 1$ ), in which case  $u_i(z_m, z)$  gives the asymptotic frequency of groups with  $i$  mutants among mutant groups. The goal is then to go from a group-level to an individual-level representation of fitness. To do so, we first left multiply

both sides of eq. (C-2) by the vector  $\mathbf{x} = (1, 2, \dots, n)$  and rearrange to get

$$\rho(z_m, z) = \frac{\mathbf{x} \cdot \mathbf{A}(z_m, z) \cdot \mathbf{u}(z_m, z)}{\mathbf{x} \cdot \mathbf{u}(z_m, z)} = \frac{\sum_{i=1}^n \sum_{j=1}^n i a_{ij}(z_m, z) u_j(z_m, z)}{\sum_{i=1}^n i u_i(z_m, z)}, \quad (\text{C-3})$$

where the denominator,  $\sum_{i=1}^n i u_i(z_m, z)$ , is the expected number of mutants in a mutant group. Second, we observe that  $\sum_{i=1}^n i a_{ij}(z_m, z)$  (which appears in the numerator of eq. C-3) corresponds to the total expected number of mutant individuals produced by a mutant group with  $j$  mutants (i.e. by all the mutant individuals in that group). As shown in [17], we can re-write this as

$$\sum_{i=1}^n i a_{ij}(z_m, z) = j w_j(z_m, z), \quad (\text{C-4})$$

where  $w_j(z_m, z)$  is the expected number of mutants produced by a mutant individual in a group with  $1 \leq j \leq n$  mutants. Plugging eq. (C-4) into eq. (C-3) allows us to write invasion fitness in terms of individual fitness:

$$\rho(z_m, z) = \sum_{k=1}^n w_k(z_m, z) q_k(z_m, z), \quad (\text{C-5})$$

where

$$q_k(z_m, z) = \frac{k u_k(z_m, z)}{\sum_{i=1}^n i u_i(z_m, z)} \quad (\text{C-6})$$

is the asymptotic probability that a mutant individual is in a group with  $k \geq 1$  mutants.

To gain further insights, we introduce the function,

$$w(z_\bullet, z_1, \dots, z_{n-1}), \quad (\text{C-7})$$

which is the fitness (i.e. expected number of descendants) of a focal individual whose trait value is  $z_\bullet \in \{z_m, z\}$  when its  $n-1$  group neighbours have traits  $z_1, \dots, z_{n-1}$  (where  $z_i \in \{z_m, z\}$  is the trait of neighbour arbitrarily indexed  $i$ ). Since groups are homogeneous and interactions within groups are random,  $w(z_\bullet, z_1, \dots, z_{n-1})$  is insensitive to permutations of its arguments  $z_1, \dots, z_{n-1}$ . In terms of eq. (C-7), the fitness of a mutant when there are  $k$  mutants in the group (which appears in invasion fitness, eq. C-5) is

$$w_k(z_m, z) = w(z_m, \underbrace{z_m, \dots, z_m}_{k-1}, \underbrace{z, \dots, z}_{n-k}), \quad (\text{C-8})$$

since there are  $k-1$  mutants and  $n-k$  residents among the neighbours of a focal mutant.

In the main text, we present results under the assumption that individual fitness can be expressed as

$$w(z_m, \underbrace{z_m, \dots, z_m}_{k-1}, \underbrace{z, \dots, z}_{n-k}) = w(z_m, \bar{z}(k), \dots, \bar{z}(k)), \quad (\text{C-9})$$

where

$$\bar{z}(k) = \frac{k-1}{n-1} z_m + \frac{n-k}{n-1} z \quad (\text{C-10})$$

is the average trait among the neighbours to a focal mutant when there are  $k$  mutants in the group. Eq. (C-9) thus posits that the focal individual whose fitness is being considered plays the field within its group: its fitness can be written as a function of its own trait and the average trait among its neighbours. For short, we denote this function by,

$$\omega(z_m, \bar{z}(k)) = w(z_m, \bar{z}(k), \dots, \bar{z}(k)). \quad (\text{C-11})$$

Plugging eq. (C-9) into eq. (C-5) we get

$$\rho(z_m, z) = \sum_{k=1}^n \omega(z_m, \bar{z}(k)) q_k(z_m, z) \quad (\text{C-12})$$

for invasion fitness, which we use to compute directional and disruptive selection (in sections C.1.3 and C.1.4). We explore these in section C.1.5 under the more general assumption that fitness is written as eq. (C-7).

### C.1.3 Directional selection

Taking the derivative of eq. (C-12), we obtain that the gradient of directional selection reduces to,

$$\begin{aligned} s(z) &= \sum_{k=1}^n \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} q_k^\circ + \sum_{k=1}^n \omega^\circ \frac{\partial q_k(z)}{\partial z_m} \\ &= \sum_{k=1}^n \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} q_k^\circ + \underbrace{\frac{\partial}{\partial z_m} \left[ \sum_{k=1}^n q_k(z_m, z) \right]}_{\substack{=1 \\ =0}} \\ &= \sum_{k=1}^n \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} q_k^\circ \end{aligned} \quad (\text{C-13})$$

where we used the fact that individual fitness under neutrality is one,  $\omega^\circ = 1$ , and that  $\sum_{k=1}^n q_k(z_m, z) = 1$  as it is a probability mass function. The derivative of individual fitness  $\omega(z_m, \bar{z}(k))$  unfolds as

$$\begin{aligned} \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} &= \frac{\partial \omega(z_\bullet, \bar{z})}{\partial z_\bullet} + \frac{\partial \bar{z}(k)}{\partial z_m} \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}} \\ &= \frac{\partial \omega(z_\bullet, \bar{z})}{\partial z_\bullet} + \frac{k-1}{n-1} \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}}, \end{aligned} \quad (\text{C-14})$$

using eq. (C-10). Substituting eq. (C-14) into eq. (C-13), we obtain

$$s(z) = \frac{\partial \omega(z_\bullet, \bar{z})}{\partial z_\bullet} + R^\circ \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}} \quad (\text{C-15})$$

where we have defined

$$R^\circ = \sum_{k=1}^n \frac{k-1}{n-1} q_k^\circ, \quad (\text{C-16})$$



as the probability a randomly sampled individual among the neighbours to a focal individual belong to the same lineage as the focal, which under neutrality corresponds to the probability that two individuals randomly sampled from the same group are identical-by-descent, i.e. to the standard coefficient of relatedness. See main text eq. (21) for interpretation of eq. (C-15).

#### C.1.4 Disruptive selection

Using the facts that  $\omega^\circ = 1$ , and that  $\sum_{k=1}^n q_k(z_m, z) = 1$ , disruptive selection calculated from eq. (C-12) can be reduced to,

$$h(z) = \underbrace{\sum_{k=1}^n \frac{\partial^2 \omega(z_m, \bar{z}(k))}{\partial z_m^2} q_k^\circ}_{=h_w(z)} + 2 \underbrace{\sum_{k=1}^n \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} \frac{\partial q_k(z_m, z)}{\partial z_m}}_{=h_r(z)}. \quad (\text{C-17})$$

Using eq. (C-10), the second-order derivative of  $\omega(z_m, \bar{z}(k))$  fitness with respect to the mutant expands as,

$$\frac{\partial^2 \omega(z_m, \bar{z}(k))}{\partial z_m^2} = \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet^2} + 2 \frac{k-1}{n-1} \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}} + \frac{(k-1)^2}{(n-1)^2} \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial \bar{z}^2}. \quad (\text{C-18})$$

so that  $h_w(z)$  in eq. (C-17) is given by

$$\sum_{k=1}^n \frac{\partial^2 \omega(z_m, \bar{z}(k))}{\partial z_m^2} q_k^\circ = \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet^2} + 2R^\circ \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}} + K^\circ \frac{\partial^2 \omega(z_\bullet, \bar{z})}{\partial \bar{z}^2} \quad (\text{C-19})$$

where we used eq. (C-16) and defined

$$K^\circ = \sum_{k=1}^n \frac{(k-1)^2}{(n-1)^2} q_k^\circ, \quad (\text{C-20})$$

as the probability that under neutrality two randomly sampled individuals (with replacement) among the neighbours to a focal individual all belong to the same lineage of the focal (i.e. are identical-by-descent). For the second term of eq. (C-17), we plug eq. (C-14) into it and obtain

$$\begin{aligned} h_r(z) &= \frac{\partial \omega(z_\bullet, \bar{z})}{\partial z_\bullet} \underbrace{\sum_{k=1}^n \frac{\partial q_k(z_m, z)}{\partial z_m}}_{=0} + \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}} \sum_{k=1}^n \frac{k-1}{n-1} \frac{\partial q_k(z_m, z)}{\partial z_m} \\ &= \frac{\partial \omega(z_\bullet, \bar{z})}{\partial \bar{z}} \frac{\partial R(z_m, z)}{\partial z_m} \end{aligned} \quad (\text{C-21})$$

where

$$\frac{\partial R(z_m, z)}{\partial z_m} = \frac{\partial}{\partial z_m} \left[ \sum_{k=1}^n \frac{k-1}{n-1} q_k(z_m, z) \right] \quad (\text{C-22})$$

is the effect of the mutant on the probability that a randomly sampled neighbour to a mutant individual is also mutant, i.e. the effect of the trait on relatedness or equivalently on mutant-mutant interactions. See main text eqs. (22)-(23) for an interpretation of these equations.

### C.1.5 Away from playing the field

Here, we highlight the effect of relaxing the assumption that individuals play the field, i.e that fitness can be written as eq. (C-9) as in [16, 18–21]. To do so, we first Taylor expand the more general function eq. (C-8) for the individual fitness of a mutant around the average trait among its neighbours (eq. C-10) up to second-order in  $\epsilon = z_m - z$  (which is sufficient for directional and disruptive selection):

$$w_k(z_m, z) = \omega(z_m, \bar{z}(k)) + \sum_{i=1}^{n-1} (z_i - \bar{z}(k)) \frac{\partial w}{\partial z_i} + \frac{1}{2} \sum_{i=1}^{n-1} (z_i - \bar{z}(k))^2 \frac{\partial^2 w}{\partial z_i^2} + \frac{1}{2} \sum_{i=1}^{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} (z_i - \bar{z}(k))(z_j - \bar{z}(k)) \frac{\partial^2 w}{\partial z_i \partial z_j} + \mathcal{O}(\epsilon^3), \quad (\text{C-23})$$

where  $w = w(z_\bullet, z_1, \dots, z_{n-1})$  for short in the derivatives. We can then use the fact that individuals within groups are homogeneous so that the effect of a trait change in any neighbour  $i$  is independent from the index  $i$ , so that  $\partial w / (\partial z_i) = \partial w / (\partial z_j)$  for all  $i, j = 1, \dots, n-1$  and similarly for higher-order derivatives. This allows us to take the derivatives out of their sums in eq. (C-23), leading us to,

$$w_k(z_m, z) = \omega(z_m, \bar{z}(k)) + \underbrace{\frac{\partial w}{\partial z_i} \sum_{i=1}^{n-1} (z_i - \bar{z}(k))}_{=0} + \frac{1}{2} \frac{\partial^2 w}{\partial z_i^2} \sum_{i=1}^{n-1} (z_i - \bar{z}(k))^2 + \frac{1}{2} \frac{\partial^2 w}{\partial z_i \partial z_j} \sum_{i=1}^{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} (z_i - \bar{z}(k))(z_j - \bar{z}(k)) + \mathcal{O}(\epsilon^3) \quad (\text{C-24})$$

$$= \omega(z_m, \bar{z}(k)) + \frac{1}{2} (n-1) \omega_2(z_m, \bar{z}(k)) \sigma(k) + \frac{1}{2} (n-1)(n-2) \omega_3(z_m, \bar{z}(k)) c(k) + \mathcal{O}(\epsilon^3),$$

where we used the definition eq. (C-10) to have  $\sum_{i=1}^{n-1} (z_i - \bar{z}(k)) = 0$ ; the shorthand notation,

$$\omega_2(z_m, \bar{z}(k)) = \left. \frac{\partial^2 w}{\partial z_i^2} \right|_{z_1 = \dots = z_{n-1} = \bar{z}(k)} \quad (\text{C-25})$$

$$\omega_3(z_m, \bar{z}(k)) = \left. \frac{\partial^2 w}{\partial z_i \partial z_j} \right|_{z_1 = \dots = z_{n-1} = \bar{z}(k)}$$

for fitness effects; as well as

$$\sigma(k) = \frac{1}{n-1} \sum_{i=1}^{n-1} (z_i - \bar{z}(k))^2 \quad (\text{C-26})$$

$$= \frac{k-1}{n-1} (z_m - \bar{z}(k))^2 + \frac{n-k}{n-1} (z - \bar{z}(k))^2$$

for the trait variance among the neighbours of a focal mutant conditional on there being  $k$  mutants in the group; and

$$\begin{aligned} c(k) &= \frac{1}{(n-1)(n-2)} \sum_{i=1}^{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} (z_i - \bar{z}(k))(z_j - \bar{z}(k)) \\ &= \frac{(k-1)(k-2)}{(n-1)(n-2)} (z_m - \bar{z}(k))^2 + 2 \frac{(k-1)(n-k)}{(n-1)(n-2)} (z_m - \bar{z}(k))(z - \bar{z}(k)) + \frac{(n-k)(n-k-1)}{(n-1)(n-2)} (z - \bar{z}(k))^2, \end{aligned} \quad (\text{C-27})$$

for the trait covariance among neighbours (also conditional on there being  $k$  mutants in the group). Plugging eq. (C-25) into eq. (C-5) we get

$$\begin{aligned} \rho(z_m, z) &= \sum_{k=1}^n \omega(z_m, \bar{z}(k)) q_k(z_m, z) \\ &\quad + \frac{1}{2}(n-1) \sum_{k=1}^n [\omega_2(z_m, \bar{z}(k))\sigma(k) + (n-2)\omega_3(z_m, \bar{z}(k))c(k)] q_k(z_m, z) \\ &\quad + \mathcal{O}(\epsilon^3) \end{aligned} \quad (\text{C-28})$$

for invasion fitness up to second-order in  $\epsilon$ .

From eq. (C-28), we see that up to first-order in  $\epsilon$  (which is sufficient to investigate directional selection),  $\rho(z_m, z) = \sum_{k=1}^n \omega(z_m, \bar{z}(k)) q_k(z_m, z)$ . Hence the selection gradient presented in the main text, although formulated in terms of the simpler fitness function  $\omega(z_m, \bar{z}(k))$ , applies for more general functions of the form of eq. (C-7). Another way to see this is that the terms that are of order  $\mathcal{O}(\epsilon^2)$  and higher in fitness eq. (C-28) do not contribute to the selection gradient as their first-order derivative with respect to  $z_m$  are proportional to  $\epsilon = z_m - z$ , which vanishes when  $z_m = z$ .

Using eq. (C-28), disruptive selection can be reduced to,

$$\begin{aligned} h(z) &= \underbrace{\sum_{k=1}^n \frac{\partial^2 \omega(z_m, \bar{z}(k))}{\partial z_m^2} q_k^\circ + \frac{1}{2}(n-1) \sum_{k=1}^n \frac{\partial^2}{\partial z_m^2} [\omega_2(z_m, \bar{z}(k))\sigma(k) + (n-2)\omega_3(z_m, \bar{z}(k))c(k)] q_k^\circ}_{=h_w(z)} \\ &\quad + 2 \underbrace{\sum_{k=1}^n \frac{\partial \omega(z_m, \bar{z}(k))}{\partial z_m} \frac{\partial q_k(z_m, z)}{\partial z_m}}_{=h_r(z)} \end{aligned} \quad (\text{C-29})$$

(where the terms that are of order  $\mathcal{O}(\epsilon^3)$  and higher in fitness eq. (C-28) can be ignored as their second-order derivative with respect to  $z_m$  are proportional to  $\epsilon = z_m - z$  and thus vanish when  $z_m = z$ ). Comparing eqs. (C-17) and (C-29), we see that  $h_w(z)$  now consists of an extra term, which we denote by

$$\Delta_w(z) = \frac{1}{2}(n-1) \sum_{k=1}^n \frac{\partial^2}{\partial z_m^2} [\omega_2(z_m, \bar{z}(k))\sigma(k) + (n-2)\omega_3(z_m, \bar{z}(k))c(k)] q_k^\circ. \quad (\text{C-30})$$

Using the chain rule for the derivatives of  $\omega_2(z_m, \bar{z}(k))\sigma(k)$  and  $\omega_3(z_m, \bar{z}(k))c(k)$  and using eqs. (C-26)-(C-27),

we obtain that this extra term reduces to

$$\Delta_w(z) = (n-2)(R^\circ - R_3^\circ)(\omega_2(z, z) - \omega_3(z, z)) \quad (\text{C-31})$$

where

$$R_3^\circ = \sum_{k=1}^n \frac{(k-1)(k-2)}{(n-1)(n-2)} q_k^\circ, \quad (\text{C-32})$$

is the probability that under neutrality three individuals randomly sampled without replacement from the same group are identical-by-descent.

Unsurprisingly,  $\Delta_w(z)$  vanishes when there are two individuals per group,  $n = 2$ . Otherwise,  $\Delta_w(z)$  will typically be non-zero when individual fitness cannot be written as eq. (C-9) (when individuals do not play the field). Nevertheless,  $\Delta_w(z)$  may in some cases be small, say of order  $\mathcal{O}(\delta)$ , so that eqs. (22)-(23) in the main text still provide a good approximation for disruptive selection even where individuals do not play the field. This is the case when for instance: (i)  $\delta = \omega_2(z, z) - \omega_3(z, z)$  is small, which from eq. C-25 we see occurs when interactions among neighbours have close to additive effects on fitness (i.e. when a second-order trait change in one neighbour has similar fitness effect as joint changes in the traits of two neighbours,  $\partial^2 w / (\partial z_i)^2 \sim \partial^2 w / (\partial z_i \partial z_j)$ ); or when (ii)  $\delta = R^\circ - R_3^\circ$  is small, which occurs when dispersal is close to zero (in which case  $R^\circ \sim 1$  and  $R_3^\circ \sim 1$ ) or 1 (in which case  $R^\circ \sim 0$  and  $R_3^\circ \sim 0$ ).

### C.1.6 Example 1

Here, we go through an example of social interactions within a group-structured populations. The main goal is to illustrate how to perform an analysis of disruptive selection under limited dispersal and how such dispersal limitation inhibits the emergence of polymorphism when cooperation has antagonistic effects among partners similar to the model of [19] (as in the snowdrift game, see Appendix A.1). We assume the population follows a Wright-Fisher life cycle (so with non-overlapping generations) with the following events happening at each generation: (i) individuals interact socially within groups, reaping material payoffs; (ii) individuals reproduce, making a large number of offspring in proportion to payoff, and then die; (iii) each offspring either disperses with probability  $d$  or remains in its natal group; (iv) finally offspring in each group compete locally for  $n$  spots, becoming the adults of the next generation.

**Fitness.** According to this life-cycle, the expected number of offspring of a focal individual with trait  $z_\bullet$  when its group neighbours have on average trait  $\bar{z}$  is

$$\omega(z_\bullet, \bar{z}) = \frac{(1-d)\pi(z_\bullet, \bar{z})}{\underbrace{(1-d)(\pi(z_\bullet, \bar{z}) + (n-1)\pi(\bar{z}, \bar{z}_{-1})) / n + d\pi(z, z)}_{\omega_p(z_\bullet, \bar{z})}} + \frac{d\pi(z_\bullet, \bar{z})}{\underbrace{\pi(z, z)}_{\omega_d(z_\bullet, \bar{z})}}, \quad (\text{C-33})$$

where  $\pi(z_\bullet, \bar{z})$  is the fecundity of such a focal individual, and where  $\bar{z}_{-1} = [z_\bullet + (n-2)\bar{z}]/(n-1)$  is the average trait among the neighbours to a neighbour of the focal. Fitness in eq. (C-33) is decomposed as the sum of two components. The first,  $\omega_p(z_\bullet, \bar{z})$ , is the expected number of offspring that establish locally, consisting of the ratio of offspring of the focal that remain in their natal group to the total number of offspring that enter the competition in that group. The second,  $\omega_d(z_\bullet, \bar{z})$ , is the expected number of offspring that establish via dispersal. Fecundity is as in the example we went through in Appendix A.1 with

$$\pi(z_\bullet, \bar{z}) = f_0 \left[ 1 - z_\bullet + B_1(z_\bullet + \bar{z}) + \frac{B_2}{2}(z_\bullet^2 + \bar{z}^2) - B_3 z_\bullet \bar{z} \right], \quad (\text{C-34})$$

so that helping has a baseline unit cost of one, while  $B_1$ ,  $B_2$  and  $B_3$  determine the benefits of helping, with in particular  $B_3$  controlling the degree of antagonism.

**Directional selection.** To determine the selection gradient as in eq. (21) requires the coefficient  $R^\circ$  of pairwise relatedness under neutrality. Such coefficient is typically computed using standard coalescent argument [e.g. 22, for textbook treatment] that we illustrate here. Suppose we are interested in the probability  $R_t^\circ$  at some generation  $t$  of sampling two haploid individuals that are identical-by-descent in a monomorphic population. Under the Wright-Fisher model of reproduction, this probability follows a recurrence,

$$R_t^\circ = (1 - m_b^\circ)^2 \left( \frac{1}{n} + \frac{n-1}{n} R_{t-1}^\circ \right), \quad (\text{C-35})$$

where  $m_b^\circ$  is the backward probability of dispersal in the resident population, i.e. the probability that a randomly sampled individual is an immigrant, which under our assumptions, is simply the probability of dispersal,

$$m_b^\circ = d. \quad (\text{C-36})$$

Accordingly,  $(1 - m_b^\circ)^2$  in eq. (C-35) is the probability of sampling two philopatric offspring (which is necessary for them to be identical-by-descent). Then, with probability  $1/n$  these individuals have the same parent in which case they are identical-by-descent, otherwise they have different parents who are themselves identical-by-descent with probability  $R_{t-1}^\circ$ . Solving eq. (C-35) for the equilibrium  $R^\circ = R_t^\circ = R_{t-1}^\circ$  yields,

$$R^\circ = \frac{(1 - m_b^\circ)^2}{1 + (n-1)[1 - (1 - m_b^\circ)^2]}, \quad (\text{C-37})$$

which in the limit of large groups and weak dispersal (i.e. as  $n \rightarrow \infty$  and  $m_b^\circ \rightarrow 0$  such as the number of immigrants  $nm_b^\circ = M$  remains constant) reads as the classical formula

$$R^\circ \rightarrow \frac{1}{1 + 2M}. \quad (\text{C-38})$$

Substituting eq. (C-33) (with eq. C-34) and eq. (C-37) into the selection leads to eq. (21)

$$s(z) = (1 - R^\circ) \frac{B_1 - 1 - z(B_3 - B_2)}{\pi(z, z)}, \quad (\text{C-39})$$

from which we obtain the singular value,

$$z^* = \frac{B_1 - 1}{B_3 - B_2}. \quad (\text{C-40})$$

Note that this singular value is independent from dispersal and the same as in a well-mixed population. This is due to our assumption that generations are non-overlapping in which case indirect fitness benefits are “cancelled out” by effects of kin competition (an increase in related offspring competing for spots in a group) [23]. The condition for convergence stability,

$$s'(z^*) = (1 - R^\circ) \frac{B_2 - B_3}{\pi(z^*, z^*)} < 0, \quad (\text{C-41})$$

is thus also the same as in a well-mixed population:  $z^*$  is convergence stable if  $B_2 - B_3 < 0$  (since  $\pi(z^*, z^*) > 0$ ). Most kin selection analyses stop here. However it may often be of interest to determine whether disruptive selection leads to social polymorphism or not. We do so in the next section.

**Disruptive selection.** Next we use eqs. (22)-(23) to compute disruptive selection. This requires  $K^\circ$ , which is the probability that under neutrality, two individuals randomly sampled with replacement among the neighbours to a random focal individual are identical-by-descent to the focal. This probability may be expressed as,

$$K^\circ = \frac{1}{n-1} R^\circ + \frac{n-2}{n-1} R_3^\circ, \quad (\text{C-42})$$

where  $R_3^\circ$  is the probability that three individuals randomly sampled without replacement from the same group are identical-by-descent, as with  $1/(n-1)$  the two individuals sampled with replacement among the neighbours to the focal are the same individual (in which case they are identical-by-descent to the focal with probability  $R^\circ$ ) and with complementary probability  $(n-2)/(n-1)$  they are different individuals (in which case they are identical-by-descent to the focal with probability  $R_3^\circ$ ). The three-way relatedness coefficient  $R_3^\circ$  is then computed using a similar argument as the one used for eq. (C-37), yielding

$$R_3^\circ = (1 - m_b^\circ)^3 \left( \frac{1}{n^2} + 3 \frac{n-1}{n^2} R^\circ + \frac{(n-1)(n-2)}{n^2} R_3^\circ \right) \quad (\text{C-43})$$

at equilibrium (we do not solve eq. C-43 explicitly for  $R_3^\circ$  but this is straightforward). One further quantity that is required to investigate disruptive selection is the effect of the trait on pairwise relatedness,  $\partial R(z_m, z) / (\partial z_m)$ . For the island model of dispersal under non-overlapping generations, this effect has been shown to be

$$\frac{\partial R(z_m, z)}{\partial z_m} = 2 \frac{R^\circ}{1 - m_b} \left[ (1 + (n-1)R^\circ) \frac{\partial \omega_p(z_\bullet, \bar{z})}{\partial z_\bullet} + (2R^\circ + (n-2)R_3^\circ) \frac{\partial \omega_p(z_\bullet, \bar{z})}{\partial \bar{z}} \right] \quad (\text{C-44})$$

([18, 19]; see [16] for overlapping generations). The above gives everything that is necessary to compute disruptive selection from eqs. (22)-(23) of the main text, which in the limit of large groups and weak dispersal tends to,

$$h(z^*) \rightarrow \frac{1}{\pi(z^*, z^*)} \frac{M}{(2M+1)(M+1)} [B_2(2M+1) - 2(B_3 - B_2)]. \quad (\text{C-45})$$

Necessary conditions for disruptive selection to occur and polymorphism to emerge are thus that  $B_2 > 0$  (as in well-mixed populations) and further that

$$M > \frac{B_3}{B_2} - \frac{3}{2}. \quad (\text{C-46})$$

There is thus a threshold number of immigrants per generation below which selection is no longer disruptive. Put differently, there is a minimum level of dispersal required for polymorphism to emerge in this model (Fig. 3A; see [18–20, 24] for similar effects). A numerical analysis of the different components of disruptive selection shows that the most negative term, and thus the one contributing most to selection being stabilising, is the interaction term  $\partial^2 \omega(z, \bar{z}) / (\partial z \partial \bar{z}) < 0$ . This indicates that interactions among relatives tend to inhibit disruptive selection and favours the evolution of equal contribution among social partners.

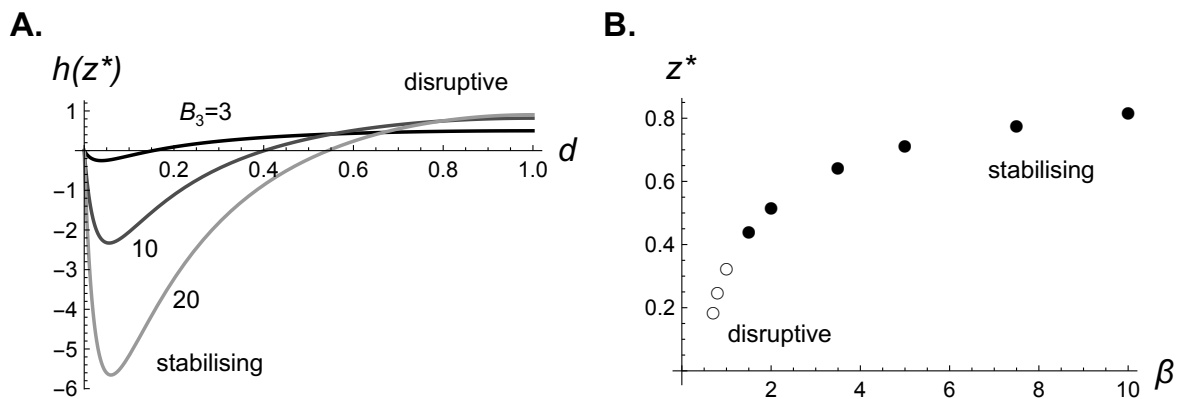


Figure 3: **Stabilising and disruptive selection in group-structured populations.** **A:** Disruptive selection against dispersal in a model of cooperation where cooperation has antagonistic effects among social partners (Appendix C.1.6 for details, with  $B_3 = 3$  in black, 10 in dark gray, 20 in light gray; other parameters:  $n = 10, B_1 = 2, B_2 = 1$ ). This indicates that disruptive selection is favoured when antagonism  $B_3$  is strong and disfavoured when dispersal  $d$  is weak. **B:** Singular level of cooperation  $z^*$  in a model where cooperation trades-off with dispersal as a function of the trade-off parameter  $\beta$  (Appendix C.1.7 for details, other parameters:  $n = 10, B_1 = 2, B_2 = 1$ ). Empty circles indicate singular strategies where selection is disruptive, i.e.  $h(z^*) > 0$ , full circles where selection is stabilising, i.e.  $h(z^*) < 0$ . This indicates that strong trade-off (when  $\beta$  is small) favour the emergence of polymorphism.

### C.1.7 Example 2

The second example for group-structured populations we go through aims to illustrate how preferential interactions among relatives can lead to polymorphism, i.e. via the term  $h_r(z^*)$ . We assume now that the dispersal

probability also depends on the evolving trait and that fitness can be expressed as,

$$\omega(z_\bullet, \bar{z}) = \frac{[1 - d(z_\bullet)] \pi(z_\bullet, \bar{z})}{\underbrace{([1 - d(z_\bullet)] \pi(z_\bullet, \bar{z}) + (n-1) [1 - d(\bar{z})] \pi(\bar{z}, \bar{z}_{-1})) / n + d(z) \pi(z, z)}_{\omega_p(z_\bullet, \bar{z})}} + \frac{d(z_\bullet) \pi(z_\bullet, \bar{z})}{\underbrace{\pi(z, z)}_{\omega_d(z_\bullet, \bar{z})}}. \quad (\text{C-47})$$

The backward probability of dispersal now also depends on the resident trait,

$$m_b^\circ = d(z). \quad (\text{C-48})$$

For social interactions, we assume that the evolving trait is a form of participation to a common good whose benefits are shared equally among all group members. Specifically we assume that fecundity now reads as

$$\pi(z_\bullet, \bar{z}) = f_0 (1 + B_1 z_0^\gamma - z_\bullet^\alpha), \quad (\text{C-49})$$

where

$$z_0 = \frac{1}{n} z_\bullet + \frac{n-1}{n} \bar{z} \quad (\text{C-50})$$

is the average contribution in the group and  $f_0 > 0$ ,  $B_1 > 0$ ,  $\gamma > 0$  and  $\alpha > 0$  are parameters controlling the fecundity benefits and cost. We assume that increased contribution to the common good also leads to decreased dispersal, due to e.g. functional trade-offs, according to

$$d(z_\bullet) = 1 - z_\bullet^\beta, \quad (\text{C-51})$$

where  $\beta > 0$  controls the shape of trade-off. We can then use the expressions for relatedness (eqs. C-37, C-42, C-43) and its perturbation (eq. C-44) given above to determine directional and disruptive selection in this model. We will not perform an exhaustive analysis of this model but a numerical exploration reveals that disruptive selection readily occurs, especially when the trade-off between cooperation and dispersal is strong (when  $\beta$  is small, Fig. 3B). This is due to the term  $h_r(z^*) = \partial\omega(z_\bullet, \bar{z})/(\partial\bar{z}) \times \partial R(z_m, z)/(\partial z_m) > 0$  as an increase in the trait simultaneously increase neighbour's fitness (due to increased participation in the common good,  $\partial\omega(z_\bullet, \bar{z})/(\partial\bar{z}) > 0$ ) and the probability that neighbours are relatives (due to decreased dispersal,  $\partial R(z_m, z)/(\partial z_m) > 0$ ; see [25–28] for similar effects but where dispersal and cooperation evolve independently).

## C.2 Heterogeneity within groups

We now turn our attention to the case where individuals within groups show heterogeneities, for instance when groups carry males and females, or individuals that have different ages, or that are in different conditions, and derive eqs. (III.A)-(IV.D) in Boxes III and IV of the main text. The analysis essentially follows the same sequence and arguments as in section C.1 but is more involved as it takes into account class variation. We assume throughout that there is a fixed number  $M$  of classes.



### C.2.1 Fitness

Here we provide an expression for invasion fitness in terms of reproductive values to capture the class structure, following [2] (their Appendix F). Our derivation also follows closely [29], perhaps clarifying a few points and highlighting a couple of assumptions that were left unsaid to derive the selection gradient. As a starting point, we introduce some notation to describe the state of a mutant group, i.e. a group with at least one mutant in one class. For a given group, we denote by  $n_i$  the number of individuals in class  $i$  in that group, and by  $k_i \leq n_i$  be the number of mutants in that class. The state of a mutant group is thus given by a vector  $\mathbf{k} = ((n_1, k_1), (n_2, k_2), \dots, (n_M, k_M))$  where at least one  $k_i$  is greater than zero. The space  $\mathcal{K}$  of all possible configurations a mutant group can be in, can thus be described as  $\mathcal{K} = \{\mathbf{k} : k_i \geq 1 \text{ for at least one } i\}$ . Then, we let  $\mathbf{A}(z_m, z)$  be a  $|\mathcal{K}| \times |\mathcal{K}|$  matrix whose element  $a_{\mathbf{k}', \mathbf{k}}(z_m, z)$  gives the expected number of groups in state  $\mathbf{k}'$  produced by a group in state  $\mathbf{k}$ . The space  $\mathcal{K}$  and associated matrix  $\mathbf{A}(z_m, z)$  may be complicated to fully characterise, but thankfully analysis of directional and disruptive selection does not require such characterisation.

As in the preceding section, invasion fitness is given by the leading eigenvalue of the matrix  $\mathbf{A}(z_m, z)$ , and satisfies

$$\rho(z_m, z) \mathbf{u}(z_m, z) = \mathbf{A}(z_m, z) \cdot \mathbf{u}(z_m, z) \quad (\text{C-52})$$

where  $\mathbf{u}$  is the right eigenvector of  $\mathbf{A}(z_m, z)$  (so of length  $|\mathcal{K}|$ ), scaled such that its entries sum to one. The entry  $u_{\mathbf{k}}(z_m, z)$  of this eigenvector is thus the asymptotic frequency of groups in state  $\mathbf{k}$  among mutant groups. The goal is to connect eq. (C-52) with an individual level measure of fitness. This is done in several steps. First, we define,

$$q_{\mathbf{k}, j}(z_m, z) = \frac{k_j u_{\mathbf{k}}(z_m, z)}{\mathbf{x} \cdot \mathbf{u}(z_m, z)} \quad (\text{C-53})$$

where  $\mathbf{x}$  is a vector of length  $|\mathcal{K}|$  whose entry  $\mathbf{k}$  gives the number of mutants in a group in such a state  $\mathbf{k}$ , i.e.

$$x_{\mathbf{k}} = \sum_{i=1}^M k_i. \quad (\text{C-54})$$

Accordingly,  $\mathbf{x} \cdot \mathbf{u}(z_m, z)$  is the expected number of mutants in a mutant group, and  $q_{\mathbf{k}, j}(z_m, z)$  is the probability that a randomly sampled mutant is in a group in state  $\mathbf{k}$  and in an individual in class  $j$ . As required, one has

$$\sum_{\mathbf{k} \in \mathcal{K}} \sum_{j=1}^M q_{\mathbf{k}, j}(z_m, z) = 1. \quad (\text{C-55})$$

We can marginalise the probability mass  $q_{\mathbf{k}, j}(z_m, z)$  to obtain the probability that a randomly sampled mutant is in an individual in class  $j$ :

$$\sum_{\mathbf{k} \in \mathcal{K}} q_{\mathbf{k}, j}(z_m, z) = q_j(z_m, z), \quad (\text{C-56})$$

which we can use to obtain

$$q_{\mathbf{k}, j}(z_m, z) = q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z), \quad (\text{C-57})$$

where  $q_{k|j}(z_m, z)$  is the conditional probability that given that a mutant is in an individual in class  $j$ , its group is in state  $\mathbf{k}$ .

Second, we define  $w_{i|j|\mathbf{k}}(z_m, z)$  as the expected number of mutant individuals in class  $i$  produced by a focal mutant individual in class  $j$ , given that its group is in state  $\mathbf{k}$ . Under neutrality (when  $z_m = z$ ), this individual fitness measure is independent of  $\mathbf{k}$  and thus reduces to  $w_{ij}^\circ$ . This allows us to define the reproductive value  $v_j^\circ$  of an individual in class  $j$ , which is such that

$$v_j^\circ = \sum_{i=1}^M v_i^\circ w_{ij}^\circ, \quad (\text{C-58})$$

and that we scale in order for,

$$\sum_{i=1}^M v_i^\circ q_i^\circ = 1, \quad (\text{C-59})$$

i.e. for the expected reproductive value of an individual that is randomly sampled from a neutral mutant lineage is one. We can then define the vector  $\mathbf{v}^\circ$  of length  $|\mathcal{K}|$  whose  $\mathbf{k} \in \mathcal{K}$  entry is

$$v_{\mathbf{k}}^\circ = \sum_{i=1}^M k_i v_i^\circ. \quad (\text{C-60})$$

Next, we left multiply eq. (C-52) by the vector  $\mathbf{v}^\circ$ , and after re-arrangement obtain,

$$\rho(z_m, z) = \frac{\mathbf{v}^\circ \cdot \mathbf{A}(z_m, z) \cdot \mathbf{u}(z_m, z)}{\mathbf{v}^\circ \cdot \mathbf{u}(z_m, z)} = \frac{1}{\mathbf{v}^\circ \cdot \mathbf{u}(z_m, z)} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{\mathbf{k}' \in \mathcal{K}} v_{\mathbf{k}'}^\circ a_{\mathbf{k}', \mathbf{k}}(z_m, z) u_{\mathbf{k}}(z_m, z) \quad (\text{C-61})$$

Using eq. (C-60) to substitute for  $v_{\mathbf{k}'}^\circ$  gives us

$$\rho(z_m, z) = \frac{1}{\mathbf{v}^\circ \cdot \mathbf{u}(z_m, z)} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{i=1}^M v_i^\circ \left[ \sum_{\mathbf{k}' \in \mathcal{K}} k'_i a_{\mathbf{k}', \mathbf{k}}(z_m, z) \right] u_{\mathbf{k}}(z_m, z), \quad (\text{C-62})$$

where the term within square brackets corresponds to the expected number of mutant individuals in class  $i$  produced by all the mutant individuals in a mutant group in state  $\mathbf{k}$ . This can therefore be equivalently written as,

$$\sum_{\mathbf{k}' \in \mathcal{K}} k'_i a_{\mathbf{k}', \mathbf{k}}(z_m, z) = \sum_{j=1}^M k_j w_{i|j|\mathbf{k}}(z_m, z). \quad (\text{C-63})$$

Plugging eq. (C-63) into eq. (C-62) then yields,

$$\rho(z_m, z) = \frac{1}{\mathbf{v}^\circ \cdot \mathbf{u}(z_m, z)} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ k_j w_{i|j|\mathbf{k}}(z_m, z) u_{\mathbf{k}}(z_m, z) \quad (\text{C-64})$$

Finally, we can multiply and divide the above by  $\mathbf{x} \cdot \mathbf{u}(z_m, z)$  and use eqs. (C-53) and (C-57) to write invasion fitness in terms of individual fitness as

$$\rho(z_m, z) = \frac{1}{V_T(z_m, z)} \sum_{\mathbf{k} \in \mathcal{K}} \sum_{i=1}^M \sum_{j=1}^M v_i^\circ w_{i|j|\mathbf{k}}(z_m, z) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z), \quad (\text{C-65})$$

where

$$V_T(z_m, z) = \frac{\mathbf{v}^\circ \cdot \mathbf{u}(z_m, z)}{\mathbf{x} \cdot \mathbf{u}(z_m, z)} = \sum_{\mathbf{k} \in \mathcal{K}} \sum_{i=1}^M v_i^\circ q_{\mathbf{k}, i}(z_m, z), \quad (\text{C-66})$$

is the expected reproductive value in a monomorphic resident population of an individual that is randomly sampled from the asymptotic mutant lineage distribution. Under neutrality, this reduces to

$$V_T^\circ = \sum_{i=1}^M v_i^\circ q_i^\circ = 1, \quad (\text{C-67})$$

thanks to normalisation eq. (C-59).

To proceed further into the analysis of selection and connect with previous results, we write the expected number of mutant individuals in class  $i$  produced by a mutant individual in class  $j$  in a group in state  $\mathbf{k}$  as an explicit function of all the traits expressed in the focal group, i.e. as a function

$$w_{ij|\mathbf{k}}(z_m, z) = w_{ij}(\underbrace{z_m}_{\text{focal}}, \underbrace{z_m, \dots, z_m}_{k_1}, \underbrace{z, \dots, z}_{n_1 - k_1}, \dots, \underbrace{z_m, \dots, z_m}_{k_{j-1}}, \underbrace{z, \dots, z}_{n_j - k_j}, \dots, \underbrace{z_m, \dots, z_m}_{k_M}, \underbrace{z, \dots, z}_{n_M - k_M}), \quad (\text{C-68})$$

whose first argument is the trait expressed by the focal mutant individual (in class  $j$ ), and the other arguments are the traits of its group neighbours: for each class  $i' \neq j$  other than of the focal, we have  $n_{i'}$  arguments for the traits of individuals in that class (composed of  $k_{i'}$  mutants and  $n_{i'} - k_{i'}$  residents); and for class  $j$  of the focal, we have  $n_j - 1$  arguments as the focal is excluded (so with  $k_j - 1$  mutant trait values and  $n_j - k_j$  residents). For simplicity, we assume that individuals “play the field”, such that we can consider that each neighbour expresses the mean trait value of its class, i.e. such that

$$w_{ij|\mathbf{k}}(z_m, z) = w_{ij}(z_m, \underbrace{\bar{z}_{1j}(\mathbf{k}), \dots, \bar{z}_{1j}(\mathbf{k})}_{n_1}, \dots, \underbrace{\bar{z}_{jj}(\mathbf{k}), \dots, \bar{z}_{jj}(\mathbf{k})}_{n_j - 1}, \dots, \underbrace{\bar{z}_{Mj}(\mathbf{k}), \dots, \bar{z}_{Mj}(\mathbf{k})}_{n_M}), \quad (\text{C-69})$$

where

$$\bar{z}_{ij}(\mathbf{k}) = \begin{cases} \frac{k_i}{n_i} z_m + \frac{n_i - k_i}{n_i} z, & i \neq j \\ \frac{k_i - 1}{n_i - 1} z_m + \frac{n_i - k_i}{n_i - 1} z, & i = j, \end{cases} \quad (\text{C-70})$$

is the average trait among the neighbours of class  $i$  to a focal individual in class  $j$  in a group in state  $\mathbf{k}$  (provided such neighbours exist). The dependence of fitness on the number of individuals in different classes within the group is thus captured via the dependence of  $\omega_{ij}$  on the mean trait in each class. For short, we denote this fitness function in terms of averages as,

$$w_{ij|\mathbf{k}}(z_m, z) = \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) \quad (\text{C-71})$$

where the vector  $\bar{\mathbf{z}}_j(\mathbf{k}) = (\bar{z}_{1j}(\mathbf{k}), \dots, \bar{z}_{Mj}(\mathbf{k}))$  collects the average trait expressed by the neighbours to a focal individual in class  $j$  in a group in state  $\mathbf{k}$ .

Substituting eq. (C-71) into eq. (C-65), we obtain

$$\rho(z_m, z) = \frac{1}{V_T(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z). \quad (\text{C-72})$$

Finally, it will be useful to use

$$\alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) = \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) - w_{ij}^\circ \quad (\text{C-73})$$

for the effect of selection on individual fitness component, which is therefore such that

$$\alpha_{ij}^\circ = w_{ij}^\circ - w_{ij}^\circ = 0. \quad (\text{C-74})$$

Plugging eq. (C-73) into eq. (C-72) gives us

$$\rho(z_m, z) = \frac{1}{V_T(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \left[ w_{ij}^\circ + \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) \right] q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z), \quad (\text{C-75})$$

which after using eqs. (C-58) and eqs. (C-66) reduces to

$$\rho(z_m, z) = 1 + \frac{1}{V_T(z_m, z)} \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z). \quad (\text{C-76})$$

This final expression for invasion fitness is then the basis for our analysis of selection that comes next.

## C.2.2 Directional selection

Taking the derivative of invasion fitness (eq. C-76) with respect to the mutant  $z_m$  and estimating it at the resident  $z$  yields,

$$\begin{aligned} s(z) &= \frac{1}{V_T^\circ} \frac{\partial}{\partial z_m} \left[ \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z) \right] \\ &+ \frac{\partial}{\partial z_m} \left[ \frac{1}{V_T(z_m, z)} \right] \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \alpha_{ij}^\circ q_{\mathbf{k}|j}^\circ q_j^\circ, \end{aligned} \quad (\text{C-77})$$

simply using the product rule. Using eq. (C-74), the second term vanishes, while the first term reduces to,

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \frac{\partial \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))}{\partial z_m} q_{\mathbf{k}|j}^\circ q_j^\circ \quad (\text{C-78})$$

(also from eq. C-74 and eq. C-67). We can then use the fact that the derivatives of  $\alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))$  and  $\omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))$  are equal at all orders (from eq. C-73), i.e. that

$$\frac{\partial^a \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))}{\partial z_m^a} = \frac{\partial^a \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))}{\partial z_m^a} \quad (\text{C-79})$$

to write the selection gradient as

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \frac{\partial \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))}{\partial z_m} q_{\mathbf{k}|j}^\circ q_j^\circ, \quad (\text{C-80})$$

which is more biologically meaningful. Next, we can use eq. (C-70) to expand the derivative of the individual fitness components as,

$$\frac{\partial \omega_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k}))}{\partial z_m} = \frac{\partial \omega_{ij}(z_\bullet, \bar{\mathbf{z}})}{\partial z_\bullet} + \frac{\partial \omega_{ij}(z_\bullet, \bar{\mathbf{z}})}{\partial \bar{z}_j} \frac{k_j - 1}{n_j - 1} + \sum_{\substack{i'=1 \\ i' \neq j}}^M \frac{\partial \omega_{ij}(z_\bullet, \bar{\mathbf{z}})}{\partial \bar{z}_{i'}} \frac{k_{i'}}{n_{i'}}, \quad (\text{C-81})$$

where the first term is the effect of a trait change in the focal individual on its own fitness, the second term of a change in the neighbours to the focal that belong to the same class  $j$ , and the rest the effect of a change in the neighbours of a different class. Substituting eq. (C-81) into eq. (C-80) then allows us to express the selection gradient as,

$$s(z) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \left[ \frac{\partial \omega_{ij}(z_\bullet, \bar{\mathbf{z}})}{\partial z_\bullet} + \sum_{i'=1}^M \frac{\partial \omega_{ij}(z_\bullet, \bar{\mathbf{z}})}{\partial \bar{z}_{i'}} R_{i'|j}^\circ \right] q_j^\circ, \quad (\text{C-82})$$

where

$$R_{i'|j}^\circ = \begin{cases} \sum_{\mathbf{k} \in \mathcal{K}} \frac{k_{i'}}{n_{i'}} q_{\mathbf{k}|j}^\circ, & i' \neq j \\ \sum_{\mathbf{k} \in \mathcal{K}} \frac{k_{i'} - 1}{n_{i'} - 1} q_{\mathbf{k}|j}^\circ, & i' = j \end{cases} \quad (\text{C-83})$$

is a class-specific coefficient of relatedness: the probability that under neutrality, an individual randomly sampled from class  $i'$  among the neighbours to a focal in class  $j$ , is identical-by-descent to the focal. See main text eq. (III.A) in Box III for interpretation of eq. (C-82).

### C.2.3 Disruptive selection

We proceed similarly to obtain disruptive selection, which using eq. (C-74) reads as

$$h(z^*) = \frac{1}{V_T^\circ} \frac{\partial^2}{\partial z_m^2} \left[ \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z) \right] + 2 \frac{\partial}{\partial z_m} \left[ \frac{1}{V_T(z_m, z)} \right] \underbrace{\frac{\partial}{\partial z_m} \left[ \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \alpha_{ij}(z_m, \bar{\mathbf{z}}_j(\mathbf{k})) q_{\mathbf{k}|j}(z_m, z) q_j(z_m, z) \right]}_{=0, \text{ when } z=z^*}, \quad (\text{C-84})$$

where the second term vanishes at a singular point (i.e. where  $s(z^*) = 0$ ). Expanding the first term with the chain rule and using eqs. (C-74) again, as well as (C-67) and (C-79), we find that disruptive selection can be expressed as the sum of three terms,

$$h(z^*) = h_w(z^*) + 2h_r(z^*) + 2h_q(z^*), \quad (\text{C-85})$$

that are given by

$$h_w(z^*) = \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \frac{\partial^2 \omega_{ij}(z_m, \bar{z}_j(\mathbf{k}))}{\partial z_m^2} q_{\mathbf{k}|j}^\circ q_j^\circ \quad (\text{C-86})$$

$$h_r(z^*) = \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \frac{\partial \omega_{ij}(z_m, \bar{z}_j(\mathbf{k}))}{\partial z_m} \frac{\partial q_{\mathbf{k}|j}(z_m, z)}{\partial z_m} q_j^\circ \quad (\text{C-87})$$

$$h_q(z^*) = \sum_{i=1}^M \sum_{j=1}^M \sum_{\mathbf{k} \in \mathcal{K}} v_i^\circ \frac{\partial \omega_{ij}(z_m, \bar{z}_j(\mathbf{k}))}{\partial z_m} q_{\mathbf{k}|j}^\circ \frac{\partial q_j(z_m, z)}{\partial z_m}, \quad (\text{C-88})$$

which we explore further below.

The first term depends on the second order derivative of  $\omega_{ij}(z_m, \bar{z})$ , which using eq. (C-70), can be expanded into,

$$\begin{aligned} \frac{\partial^2 \omega_{ij}(z_m, \bar{z})}{\partial z_m^2} &= \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet^2} + 2 \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}_j} \frac{k_j - 1}{n_j - 1} + 2 \sum_{\substack{i'=1 \\ i' \neq j}}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}_{i'}} \frac{k_{i'}}{n_{i'}} \\ &+ 2 \sum_{\substack{i'=1 \\ i' \neq j}}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_j \partial \bar{z}_{i'}} \frac{k_j - 1}{n_j - 1} \frac{k_{i'}}{n_{i'}} + \sum_{\substack{i'=1 \\ i' \neq j}}^M \sum_{\substack{i''=1 \\ i'' \neq i' \neq j}}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'} \partial \bar{z}_{i''}} \frac{k_{i''}}{n_{i''}} \frac{k_{i'}}{n_{i'}} \\ &+ \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_j^2} \frac{(k_j - 1)^2}{(n_j - 1)^2} + \sum_{\substack{i'=1 \\ i' \neq j}}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'}^2} \frac{k_{i'}^2}{n_{i'}^2}. \end{aligned} \quad (\text{C-89})$$

Plugged into eq. (C-86), we can then write the first term of disruptive selection as,

$$h_w(z^*) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \left[ \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet^2} + 2 \sum_{i'=1}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet \partial \bar{z}_{i'}} R_{i'|j}^\circ + \sum_{i'=1}^M \sum_{i''=1}^M \frac{\partial^2 \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'} \partial \bar{z}_{i''}} K_{i', i''|j}^\circ \right] q_j^\circ, \quad (\text{C-90})$$

where we used eq. (C-83) and defined

$$K_{i', i''|j}^\circ = \begin{cases} \sum_{\mathbf{k} \in \mathcal{K}} \frac{(k_j - 1)^2}{(n_j - 1)^2} q_{\mathbf{k}|j}^\circ, & i' = i'' = j \\ \sum_{\mathbf{k} \in \mathcal{K}} \frac{k_{i'}^2}{n_{i'}^2} q_{\mathbf{k}|j}^\circ, & i' = i'' \neq j \\ \sum_{\mathbf{k} \in \mathcal{K}} \frac{k_{i'} k_j - 1}{n_{i'} n_j - 1} q_{\mathbf{k}|j}^\circ, & i' \neq i'' = j \\ \sum_{\mathbf{k} \in \mathcal{K}} \frac{k_{i'} k_{i''}}{n_{i'} n_{i''}} q_{\mathbf{k}|j}^\circ, & i' \neq i'' \neq j \neq i' \end{cases}, \quad (\text{C-91})$$

which is the probability that under neutrality two individuals randomly sampled with replacement, one from class  $i'$  and the other from class  $i''$ , among the neighbours to a focal individual in class  $j$  all belong to the same lineage of the focal (i.e. are identical-by-descent). For the second second term participating to disruptive

selection, we substitute eq. (C-81) into eq. (C-87) to obtain,

$$\begin{aligned}
h_r(z^*) &= \sum_{i=1}^M v_i^\circ \sum_{j=1}^M \left[ \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet} \underbrace{\sum_{k \in \mathcal{K}} \frac{\partial q_{k|j}(z_m, z)}{\partial z_m}}_{=0} \right. \\
&\quad \left. + \sum_{k \in \mathcal{K}} \left( \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_j} \frac{k_j - 1}{n_j - 1} + \sum_{\substack{i'=1 \\ i' \neq j}}^M \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'}} \frac{k_{i'}}{n_{i'}} \right) \frac{\partial q_{k|j}(z_m, z)}{\partial z_m} \right] q_j^\circ \\
&= \sum_{i=1}^M \sum_{j=1}^M \sum_{i'=1}^M v_i^\circ \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'}} \frac{\partial R_{i'|j}(z_m, z)}{\partial z_m} q_j^\circ
\end{aligned} \tag{C-92}$$

where we used the fact that  $\sum_{k \in \mathcal{K}} q_{k|j}(z_m, z) = 1$  so that the first line vanishes, and defined

$$\frac{\partial R_{i'|j}(z_m, z)}{\partial z_m} = \begin{cases} \frac{\partial}{\partial z_m} \left[ \sum_{k \in \mathcal{K}} \frac{k_{i'}}{n_{i'}} q_{k|j}(z_m, z) \right], & i' \neq j \\ \frac{\partial}{\partial z_m} \left[ \sum_{k \in \mathcal{K}} \frac{k_{i'} - 1}{n_{i'} - 1} q_{k|j}(z_m, z) \right], & i' = j \end{cases} \tag{C-93}$$

for the effect of the mutant on the probability that a randomly sampled neighbour in class  $i'$  to a mutant individual in class  $j$  is also mutant. Finally, substituting eq. (C-81) into eq. (C-88) and using eq. (C-83), we get that the last term participating to disruptive selection is,

$$h_q(z^*) = \sum_{i=1}^M \sum_{j=1}^M v_i^\circ \left[ \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial z_\bullet} + \sum_{i'=1}^M \frac{\partial \omega_{ij}(z_\bullet, \bar{z})}{\partial \bar{z}_{i'}} R_{i'|j}^\circ \right] \frac{\partial q_j(z_m, z)}{\partial z_m}. \tag{C-94}$$

See main text eqs. (IV.A)-(IV.D) in Box IV for interpretation of these terms.

## References

- [1] Takada T, Nakajima H. 1998 Theorems on the invasion process in stage-structured populations with density-dependent dynamics. *Journal of Mathematical Biology* **36**, 497D514.
- [2] Lehmann L, Mullon C, Akçay E, Van Cleve J. 2016 Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution* **70**, 1689–1702.
- [3] Taylor PD. 1990 Allele-frequency change in a class-structured population. *The American Naturalist* **135**, 95–106.
- [4] Caswell H. 2001 *Matrix population models: construction, analysis, and interpretation*. Sinauer associates, Inc., Sunderland, MA.
- [5] Fisher R. 1930 *The genetical theory of natural selection*. Clarendon Press.

- [6] Charlesworth B. 1994 *Evolution in age-structured populations* vol. 2. Cambridge University Press Cambridge.
- [7] Durinx M, Meszéna G et al.. 2008 Adaptive dynamics for physiologically structured population models. *Journal of mathematical biology* **56**, 673–742.
- [8] Avila P, Priklopil T, Lehmann L. 2021 Hamilton's rule, gradual evolution, and the optimal (feedback) control of phenotypically plastic traits. *Journal of Theoretical Biology* **526**, 110602.
- [9] Lehmann L. 2022 Hamilton's rule, the evolution of behavior rules and the wizardry of control theory. *Journal of Theoretical Biology* **555**, 111282.
- [10] Liberzon D. 2011 *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press.
- [11] Athans M, Falb PL. 2013 *Optimal control: an introduction to the theory and its applications*. Courier Corporation.
- [12] Metz JAJ, Staňková K, Johansson J. 2016 The canonical equation of adaptive dynamics for life histories: from fitness-returns to selection gradients and Pontryagin's maximum principle. *Journal of Mathematical Biology* **72**, 1125–1152.
- [13] Day T, Taylor PD. 2000 A generalization of Pontryagin's maximum principle for dynamic evolutionary games among relatives. *Theoretical Population Biology* **57**, 339–356.
- [14] Hille E, Phillips RS. 1957 *Functional analysis and semi-groups*. Providence, Rhode Island: American Mathematical Society.
- [15] Anton H, Rorres C. 2013 *Elementary linear algebra: applications version*. John Wiley & Sons.
- [16] Mullon C, Keller L, Lehmann L. 2016 Evolutionary stability of jointly evolving traits in subdivided populations. *The American Naturalist* **188**.
- [17] Mullon C, Lehmann L. 2014 The robustness of the weak selection approximation for the evolution of altruism against strong selection. *Journal of Evolutionary Biology*.
- [18] Ajar. 2003 Analysis of disruptive selection in subdivided populations.. *BMC Evolutionary Biology* **3**, 22.
- [19] Wakano JY, Lehmann L. 2014 Evolutionary branching in deme-structured populations. *Journal of Theoretical Biology* **351**, 83–95.
- [20] Mullon C, Lehmann L. 2019 An evolutionary quantitative genetics model for phenotypic (co)variances under limited dispersal, with an application to socially synergistic traits. *Evolution* **73**, 1695–1728.
- [21] Ohtsuki H, Rueffler C, Wakano JY, Parvinen K, Lehmann L. 2020 The components of directional and disruptive selection in heterogeneous group-structured populations. *Journal of Theoretical Biology* **507**.



- [22] Rousset F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, NJ: Princeton University Press.
- [23] Taylor PD. 1992 Inclusive Fitness in a Homogeneous Environment. *Proceedings of the Royal Society B: Biological Sciences* **249**, 299–302.
- [24] Day T. 2001 Population structure inhibits evolutionary diversification under competition for resources. *Genetica* **112-113**, 71–86.
- [25] Koella JC. 2000 The spatial spread of altruism versus the evolutionary response of egoists. *Proceedings of the Royal Society of London Series B-Biological Sciences* **267**, 1979–1985.
- [26] Powers ST, Penn AS, Watson RA. 2011 The concurrent evolution of cooperation and the population structures that support it. *Evolution* **65**, 1527–1543.
- [27] Parvinen K. 2013 Joint evolution of altruistic cooperation and dispersal in a metapopulation of small local populations. *Theoretical Population Biology* **85**, 12–19.
- [28] Mullan C, Keller L, Lehmann L. 2018 Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. *Nature Ecology and Evolution* **2**, 132–140.
- [29] Taylor PD, Frank SA. 1996 How to make a kin selection model. *Journal of Theoretical Biology* **180**, 27–37.