**Major article**

# Hierarchical variation in phenotypic flexibility across timescales and associated survival selection shape the dynamics of partial seasonal migration

**Running head:** Phenotypic flexibility and selection

**Keywords**: disruptive selection, multistate capture-recapture, partial migration, phenotypic plasticity, temporally varying selection, threshold trait

**Number of words in main text**: 7332

**Manuscript elements**: Abstract, Introduction, Material and Methods, Results, Discussion, Acknowledgements, Statement of Authorship, Data and Code Accessibility, Literature cited, Figure 1 to 7, Online Supporting Material S1 to S4.

Paul Acker[1,2]*, Francis Daunt[3], Sarah Wanless[3], Sarah J. Burthe[3]*, Mark A. Newell[3], Michael P. Harris[3], Carrie Gunn[3], Robert Swann[4], Ana Payo-Payo[2], and Jane M. Reid[1,2]

[1] Centre for Biodiversity Dynamics, Institutt for Biologi, NTNU, Norway

[2] School of Biological Sciences, University of Aberdeen, Scotland

[3] UK Centre for Ecology & Hydrology, Bush Estate, Penicuik, Scotland

[4] Highland Ringing Group, Tain, Scotland

*Corresponding author: paul.acker@ntnu.no

# Abstract

Population responses to environmental variation ultimately depend on within-individual and among-individual variation in labile phenotypic traits that affect fitness, and resulting episodes of selection. Yet, complex patterns of individual phenotypic variation arising within and between time periods, and associated variation in selection, have not been fully conceptualised or quantified. We highlight how structured patterns of phenotypic variation in dichotomous threshold traits can theoretically arise and experience varying forms of selection, shaping overall phenotypic dynamics. We then fit novel multistate models to ten years of band-resighting data from European shags to quantify phenotypic variation and selection in a key threshold trait underlying spatio-seasonal population dynamics: seasonal migration versus residence. First, we demonstrate substantial among-individual variation alongside substantial between-year individual repeatability in within-year phenotypic variation ('flexibility'), with weak sexual dimorphism. Second, we demonstrate that between-year individual variation in within-year phenotypes ('supraflexibility') is structured and directional, consistent with the threshold trait model. Third, we demonstrate strong survival selection on within-year phenotypes, and hence on flexibility, that varies across years and sexes, including episodes of disruptive selection representing costs of flexibility. By quantitatively combining these results, we show how supraflexibility and survival selection on migratory flexibility jointly shape population-wide phenotypic dynamics of seasonal movement.

# Introduction

Population responses to environmental variation and change ultimately arise through within-individual and among-individual variation in phenotypic expression of key environmentally-sensitive traits, and resulting variation in fitness components and consequent selection (Nussey et al. 2007; Gienapp et al. 2008; Chevin et al. 2010; Fox et al. 2019). Quantifying temporal dynamics of individual phenotypes within and between consecutive time periods, and quantifying associated dynamics of selection, is consequently necessary to understand and predict joint phenotypic, population dynamic and micro-evolutionary outcomes across temporal scales (Hendry 2016; Araya-Ajoy et al. 2015; Arnold et al. 2019; de Villemereuil et al. 2020).

Rapid environmentally-induced changes in phenotypes occurring on short (within-generation) timescales typically primarily reflect within-individual responses, representing phenotypic plasticity of labile traits (Sih et al. 2011; Gabriel et al. 2005; Charmantier et al. 2008; Chevin et al. 2010). In wild settings where all pertinent axes of underlying environmental variation and resulting multi-dimensional reaction norms are hard to explicitly quantify, such plastic responses can be usefully summarised as patterns and magnitudes of observed within-individual phenotypic variation (here termed 'phenotypic flexibility' or simply 'flexibility'; Piersma and Drent 2003; Glossary and summary of key terms in Online Supplemental Material, 'OSM', S1). Hierarchical levels of flexibility could then arise across timescales; individuals could be phenotypically invariable or variable within time periods (e.g. within years), while their within-period sequence of phenotypes could vary between periods (e.g. between years). Such expression of different levels of phenotypic flexibility arising across the hierarchy of timescales could then experience episodes of selection, which could be consistent or divergent, and could differ between groups of individuals (e.g. sexes), and/or vary in relation to environmental conditions. However, despite their central roles in shaping population responses, key aspects of expected and realised hierarchical patterns of phenotypic flexibility across nested temporal scales, and associated patterns of selection, have not yet been fully conceptualised or quantified for key traits that directly link phenotypic and population dynamic outcomes. Here, we provide a conceptual framework to rationalise hierarchical patterns of phenotypic flexibility and selection on flexibility concerning dichotomous threshold traits. We apply this framework to estimate key effects for the ecologically critical threshold trait of seasonal migration versus residence.

Recent studies focussing primarily on traits that are continuously distributed on observed phenotypic scales have established that reaction norm slopes defining phenotypic plasticity can vary among individuals within populations (Nussey et al. 2007; Brommer 2013; Dingemanse and Wolf 2013; Hendry 2016). Some individuals consequently show more phenotypic variation than others across a given time period or axis of environmental variation. Further, such among-individual differences may be repeatable (Araya-Ajoy and Dingemanse 2017; Mitchell and Biro 2017; Cornwell et al. 2019; Strickland et al. 2021), meaning that individuals show consistent degrees of relative phenotypic variation. Nonetheless, reported repeatabilities of individual reaction norm slopes of ~0.4–0.5 (Araya-Ajoy and Dingemanse 2017; Mitchell and Biro 2017; Strickland et al. 2021) also imply potential for substantial within-individual variation in phenotypic variation. Individuals could consequently vary in the extent to which they change their phenotype between time periods or environmental axes. Yet, key aspects of the potential for structured forms of among-individual variation in within- versus between-period phenotypic variation, and associated forms of selection, still remain to be conceptualised and quantified.

In particular, numerous labile behavioural and life-history traits that link environmental changes with population outcomes are expressed as discrete alternative phenotypes rather than as continuously distributed phenotypes (e.g. dominant versus subordinate behaviour, breeding versus skipping, mate fidelity versus divorce, or seasonal migration versus residence; Snell-Rood et al. 2018; Reid and Acker 2022). Such traits can often be appropriately conceptualised as quantitative genetic ‘threshold traits’. Here, an underlying continuously distributed latent variable termed ‘liability’, which can comprise combinations of genetic and environmental effects, translates into expression of discrete alternative phenotypes X or Y when below versus above some threshold value (fig. 1A; Gianola 1982; Roff 1996; Lynch and Walsh 1998 Ch.25). Variation occurring on the phenotypic scale can then be decoupled from reaction norm slopes defining liability-scale plasticity, and instead reflect variation in liability-scale reaction norm elevations. There can consequently be among-individual variation in observed phenotypic flexibility without any among-individual variation in liability-scale reaction norm slopes, and conversely there can be among-individual variation in liability-scale reaction norm slopes that does not result in among-individual variation in phenotypic flexibility (Reid and Acker 2022). Such threshold traits could consequently generate distinctive and biologically important patterns of phenotypic variation within and among individuals, within and among temporal scales, that could in turn experience varying selection (fig. 1, 2).

Specifically, within any focal time period over which selection could act, individuals might inflexibly express phenotype X or phenotype Y, or flexibly switch between the two (fig. 1A,B). The set of possible within-period phenotypic sequences (hereafter ‘tactics’) constitutes a continuum of within-individual expression of X versus Y (summarized as ‘full-X’, ‘mixed-XY’ and ‘full-Y’), reflecting underlying individual liability elevations (fig. 1B,C, 2A). Then, between any two consecutive time periods, individuals could either retain the same tactic, or switch tactic, and thereby potentially alter their degree of flexibility (fig. 2A). We term such between-period variation 'supraflexibility' to stress the hierarchy of phenotypic expression, and hence of flexibility (OSM S1). Such hierarchical timeframes are necessarily relative, meaning that the relevant duration of periods through which flexibility and supraflexibility are quantified will be system-, trait- and question-specific. Given the threshold trait model, the distance of an individual's initial liability from the threshold (i.e. the elevation) will affect whether or not subsequent environmental variation will cause its liability to cross the threshold, and thereby cause phenotypic change (fig. 1B; Reid and Acker 2022). Accordingly, flexibility and supraflexibility are likely to be structured and directional, such that some transitions occur more frequently than others. Specifically, full-X and full-Y individuals should be more likely to switch to mixed-XY than to full-Y and full-X respectively. Supraflexibility should also intrinsically covary with flexibility, such that individuals that are more flexible within periods (i.e. mixed-XY) could also on average be more likely to subsequently switch to full-X or full-Y than vice versa (fig. 2A). Flexibility and supraflexibility in phenotypic expression of a labile threshold trait can therefore represent different manifestations of the same underlying reaction norm enacted on the latent liability scale, but occurring on hierarchical timeframes (e.g. Reid and Acker 2022).

Selection could then act on flexibility and/or on supraflexibility if individual tactics affect relative survival and/or reproduction (fig. 2B,C). For example, inflexible individuals that express full-X or full-Y during a given period could potentially have similar survival probabilities, while flexible individuals that express mixed-XY could have higher or lower survival probability. These scenarios could represent benefits or costs of phenotypic flexibility, and generate stabilizing or disruptive selection along the X–Y continuum (fig. 2B). Yet, in other periods, environmental conditions could potentially mean that survival is higher for full-Y than full-X and intermediate for mixed-XY, generating directional selection for Y on the X–Y continuum (or, conversely, directional selection for X; fig. 2B). Here, supraflexible individuals that switched from full-X to mixed-XY, or from mixed-XY to full-Y, between two such consecutive periods would then have higher overall survival probability than individuals that did not switch, effectively generating selection for supraflexibility (fig. 2C). Yet, such effects of sequences of selection on flexibility in generating selection on supraflexibility could be further compounded by direct selection against supraflexibility, for example manifested as lower survival of individuals that is directly caused by switching tactics between periods. The overall shape and net strength of selection acting on supraflexibility could then vary across time (e.g. fig. 2B,C), driving complex short-term phenotypic dynamics within and across time periods, as well as potentially shaping longer-term evolutionary change.

Specifically, episodes of survival selection could directly cause additional within-generation shifts in population-wide frequencies of phenotypic tactics, and hence of flexible and inflexible individuals. The magnitude of such shifts will depend on among-individual variance in relative fitness (i.e. the opportunity for selection; Arnold and Wade 1984), which in turn depends on relative tactic frequencies prior to selection. Resulting within-generation shifts could persist across time periods if tactics are highly repeatable, or could be quickly mitigated or even reoriented, depending on the magnitude and directionality of supraflexibility that non-randomly alters tactic frequencies among surviving individuals. Moreover, episodes of sex-specific selection could directly generate or accentuate sex biases in tactic frequencies, reflecting sexual dimorphism in underlying liability and implying sexual dimorphism in flexibility and/or supraflexibility. Thus, while sex-specific selection in continuously distributed phenotypic traits can generally shape evolutionary change and drive evolution of sexual dimorphism (Lande 1980; Kruuk et al. 2008; Cox and Calsbeek 2009), such outcomes can be further affected by the properties of threshold traits (Reid and Acker 2022). Specifically, if there is sexual dimorphism in phenotypic expression of the threshold trait, there will be intrinsic sexual dimorphism in intensity of selection (Falconer and Mackay 1996). Overall, therefore, understanding and predicting phenotypic and population outcomes requires quantifying sex-specific frequencies of flexible and inflexible tactics, and associated patterns of supraflexibility and temporal variation in survival selection, regarding key environmentally-sensitive dichotomous traits that drive population dynamics. But, such effects have not previously been explicitly quantified, or subsequently combined to evaluate relative effects on population-level phenotypic variation.

One ecologically critical threshold trait is seasonal migration (i.e. reversible movement across seasons, hereafter ‘migration’; Dingle 1996; Newton 2008) versus year-round residence in populations with facultative migration. Resulting ‘partial migration’, where sympatric-breeding individuals can express alternative resident and migrant phenotypes (hereafter ‘R’ and ‘M’) through non-breeding seasons, occurs in diverse taxa including fish, mammals, birds, reptiles, amphibians (Lundberg 1988; Chapman et al. 2011; Reid et al. 2018). Because R and M individuals are by definition seasonally spatially segregated, local environmental variation can cause episodes of strong survival selection, notably during extreme climatic events (‘ECEs’) that affect some locations more than others (Acker et al. 2021a). Further, various annual migratory tactics may be expressed, such that some individuals inflexibly express R or M throughout the non-breeding season ('full-R' or 'full-M'), while others flexibly switch between R and M ('mixed-RM'), for example manifested as late departure from breeding areas (Cagnacci et al. 2011; Fudickar et al. 2013; Reid et al. 2020). Any change in phenotypic expression along the R-M continuum, and hence in population-wide frequencies of full-R, mixed-RM, and full-M, will then directly alter spatio-seasonal population dynamics (Reid et al. 2018). However, even though variation in expression of migration versus residence is currently attracting considerable attention and has long been conceptualised as a threshold trait (Berthold 1988; Pulido et al. 1996; Dodson et al. 2013), the hierarchical structure of sex-specific flexibility and supraflexibility, and associated selection, have not been fully quantified.

To achieve these ambitions, we fitted a novel multistate capture-recapture model to ten years of large-scale year-round capture-resighting data from individually marked adult European shags (*Gulosus aristotelis*, hereafter 'shags'). Our model represents individual phenotypic variation along the R–M continuum as a dynamic mixture of annual tactics, allowing us to quantify four sets of effects. First, we estimated year- and sex-specific tactic frequencies, and thereby quantified the degree of among-individual variation in within-year phenotypic flexibility in migration versus residence. Second, we estimated probabilities of all possible forms of tactic switching between consecutive years, and hence quantified sex-specific tactic repeatability and structured supraflexibility. Third, we estimated tactic-dependent survival probabilities, hence quantified temporal variation in the form of sex-specific survival selection on migratory tactics, and hence on migratory flexibility. Finally, we combined all these estimates to quantify relative effects of supraflexibility and survival selection on flexibility, alongside entry of new adults into the analysed population, on population-wide changes in tactic frequencies (including flexibility) from each year to the next. We thereby illustrate how phenotypic dynamics in a key trait that directly affects seasonal population distributions can be shaped by combinations of sex-specific flexibility, supraflexibility, and selection on flexibility in the context of among-year environmental variation. We discuss these outcomes in the context of general theoretical expectations for phenotypic and micro-evolutionary dynamics of plastic threshold traits.

## Material and methods

### *Study system and data collection*

Quantifying variation in phenotypic flexibility and supraflexibility of migration versus residence, and associated survival selection, requires repeated measurement of locations and survival within and across multiple years in numerous individuals. We collected the required data using visual field resightings during 2009–2019 in partially-migratory shags that breed on Isle of May National Nature Reserve (hereafter 'IoM'), Scotland (56°11′N, 2°33′W; Daunt et al. 2014; Grist et al. 2014; Keogan et al. 2021). Shags are seabirds that must return to shore every day to dry their partially-wettable plumage (Harris and Swann 2002), and migrants in our focal population typically winter 100-500 km from IoM. Marked individuals can consequently be resighted onshore within the UK throughout each year, allowing direct observation of individuals' winter locations (Grist et al. 2014; Acker et al. 2021a), and hence allowing inference on resident versus migrant phenotypes and resulting annual migratory tactics.

To generate a population of marked adults, ≥80% of chicks hatched on IoM since 1997 were marked before fledging with an individually-coded colour band (field readable up to ~150m with a telescope) and an inscribed metal band. Since ~90% of individuals that survived to adulthood recruited locally (median age 3 years; Barlow et al. 2013), by 2009 most breeding adults were banded. Remaining unbanded adults, including incoming immigrants, were caught and banded during breeding as far as feasible. Every non-breeding season (early September to late February, hereafter ‘winter’) from 2009-10 to 2018-19, resighting surveys were carried out approximately biweekly within the residency area (IoM and adjacent day roosts) and at key alternative wintering locations across eastern Scotland, with additional resightings collected elsewhere (Grist et al. 2014; Acker et al. 2021a; summarised in OSM S2). During subsequent breeding seasons (April–June 2010 to 2019), nests and roost sites on IoM were surveyed at least twice weekly to identify colour-banded adults, resulting in very high breeding season resighting probabilities (mean 0.95, range 0.90-0.98) facilitating inferences on annual survival (Acker et al. 2021a). Sexes of most adults were determined through vocalizations and/or genotyping (Acker et al. 2021a).

### *Model design*

In general, unbiased estimation of phenotypic variation and associated selection requires analyses that account for processes leading to non-random missing data due to inevitable imperfections in abilities to measure or assign individual phenotypes and/or associated fitness (Gimenez et al. 2008; Hadfield 2008; Nakagawa and Freckleton 2008). Highly non-random missing data inevitably arise when the focal trait is migration versus residence inferred from field observations. Here, marked individuals can be missed or move outside surveyed areas, causing resighting failure that varies according to phenotype given spatio-temporal variation in observational efficiency. Missing data following the last resighting of an individual could also reflect mortality, confounding estimates of phenotype-dependent resighting and survival probabilities (and hence of selection). Further, our current focal phenotypic tactics (namely season-long full-R, full-M and mixed-RM) can never be instantaneously observed, but must be inferred from sequences of observations, which are often incomplete due to resighting failure and/or within-winter mortality. Consequently, (generalized) linear mixed models that are widely used to estimate within- and among-individual phenotypic variation in readily observed traits (e.g. Araya-Ajoy et al. 2015) cannot be directly applied to estimate migratory flexibility, supraflexibility and associated selection, at least without an observation model to account for non-random missing data.

Hence, to answer our current questions, we devised a novel full-annual-cycle multistate model that uses capture-recapture methods to enable robust probabilistic inference on variable phenotypic expression of residence versus migration and associated survival from large-scale mark-recapture data given observation failure that is non-random with respect to both tactic and survival. Specifically, our model represents the process of within-year transitions between states defined by current individual location, parameterised by movement and survival probabilities, and overlaid by the observation process parameterised by resighting probabilities. By including finite individual heterogeneity defined through constraints on transitions between resident and migrant states within years, we explicitly modelled within- and among-individual phenotypic variation (and hence variation in flexibility and supraflexibility) as a dynamic mixture of annual migratory tactics parameterised by initial tactic probabilities and tactic switching probabilities (fig. 3).

To maximise use of available data, we divided the annual cycle into five resighting occasions comprising the breeding season (occasion 1) and four occasions through the subsequent winter (occasions 2–5, fig. 3; OSM S2). We defined states representing six spatially-distinct areas: the ‘residency area’ (IoM area), four observed migrant areas (defined based on geography and survey frequency), and one unobserved migrant area comprising unsurveyed migrant destinations (i.e. a ‘ghost area’; OSM S2). This spatial structure allowed us to account for spatio-temporal variation in resighting probability and hence minimise related bias (Acker et al. 2021a). Accordingly, alive individuals can be resighted where they are located, or not resighted, according to sex×occasion×year×area-dependent (‘×’ denotes interacting effects) resighting probability ($p$, which is zero in the ghost area; OSM S3). This requires that migrants are spatially distributed according to probabilities of moving to a specific migratory area conditional on departure (δ, occasion×year×area-dependent) and of moving between migratory areas conditional on not returning to the residency area (σ, assumed constant across occasions and years; Acker et al. 2021a; OSM S3). This spatial division of migrant destinations does not affect the current focus on dichotomous migrant versus resident phenotypes (Chapman et al. 2011; Pulido 2011; Dodson et al. 2013), defined based on departure from and return to the residency area (Acker et al. 2021a).

Specifically, every breeding season (occasion 1, when all individuals are located in the residency area), new individuals enter the dataset as adults and are envisaged to undertake a full-R, mixed-RM or full-M tactic through the subsequent winter (i.e. occasions 2-5; fig. 3), according to sex×year-dependent initial tactic probabilities ($\pi$; OSM S3). Each tactic is characterised by specific constraints on departure from and return to the residency area between winter occasions. Full-R individuals remain in the residency area through all annual occasions. Full-M individuals depart from the residency area between occasion 1 and 2 and do not return until subsequent occasion 1 (fig. 3). Meanwhile, mixed-RM individuals enact a phenotypic sequence that differs from both full-R and full-M (for example, by departing after occasion 2 or 3, representing within-year phenotypic flexibility; fig. 3, OSM S3). To account for potential variation in migration timings, phenotypic variation occurs in mixed-RM following probabilities of departing from ($\varepsilon$) and returning to ($\tau$) the residency area, which are sex×occasion×year×area-dependent. Previous analyses of two years of data from a different partially migratory shag population showed that division into the three full-R, full-M and mixed-RM migratory tactics provides a good representation of overall phenotypic variation (Reid et al. 2020; OSM S2). Following these previous analyses and inspection of our current dataset, occasion time windows (particularly occasion 2) were defined to appropriately distinguish the three tactics (OSM S2).

Between consecutive years, individuals either repeat the same tactic or switch to another tactic, following sex×year×tactic-dependent switching probabilities ($\kappa$). The tactic mixture is consequently dynamic, and allows for structured supraflexibility of migration versus residence (fig. 3, OSM S3). Mortality occurs between occasions according to sex×tactic×occasion×year-dependent survival probabilities ($\phi$), that are multiplied across occasions within years to yield sex×tactic×year-dependent annual survival probabilities ($\Phi$; OSM S3). Since there is virtually no breeding dispersal from IoM (Aebischer 1995; Barlow et al. 2013), these survival probabilities represent true survival probabilities for focal adults (Acker et al. 2021a).

Overall, this model fulfils our current objectives of estimating sex-specific initial tactic probabilities and tactic switching probabilities and thereby estimating variation in flexibility and supraflexibility of migration versus residence (fig. 2A) in each year and sex, and of estimating associated survival probabilities and hence survival selection on flexibility (fig. 2B). Our current model does not estimate survival selection on supraflexibility (fig. 2C), which requires major future developments and data (see *Discussion*). Since previous and exploratory analyses did not reveal major age effects on tactic frequencies (Reid et al. 2020) or survival probabilities (Acker et al. 2021a) in adults, and incorporating age effects within our model would entail substantial extra complexity, such effects were not further considered in current analyses.

### *Dataset and model analysis*

We fitted the model to capture-resighting histories of 2304 adult shags (1185 females and 1119 males), compiled from 48,413 year-round field sightings (OSM S2). Each history comprises a sequence of observation events indicating whether and where the individual was resighted in each occasion, starting from their first breeding attempt on IoM during 2009–2018 and ending in breeding season 2019.

Previous cross-sectional analyses spanning 2009–2018 showed that individuals that were resident versus migrant at any occasion (i.e. not accounting for annual tactics) typically had similarly high survival probabilities. However, ECEs comprising extreme late-winter storms caused substantial mortality in winters 2012–13, 2013–14 and 2017–18, and also caused episodes of strong survival selection against current residents in winters 2012–13 and 2017–18 (Acker et al. 2021a). These ECEs now provide opportunities to assess whether frequencies of annual migratory tactics (and hence flexibility), between-year switching (i.e. supraflexibility) and/or selection varied with known major environmental perturbations. Accordingly, in our current analyses, $\Phi$ for each tactic was considered to be constant across all non-ECE years, but allowed to differ in each ECE year (2012-13, 2013-14, 2017-18). Direct ECE effects on expression of migratory tactics in 2017-18 cannot be detected by our current model because the ECE occurred in March (i.e. after the last specified winter occasion 5), but tactic-specific survival probabilities can still be estimated.

To estimate model parameters, we used Stan, a probabilistic programming language for Bayesian inference, called from R through package *rstan* (Carpenter et al. 2017; R core team 2020). For all parameters, we used priors that were uniform over the natural range of probabilities or probability simplexes (OSM S3). Posterior predictive checks (Gelman et al. 1996) devised for full-annual-cycle multistate capture-recapture analyses (Acker et al. 2021a) indicated good overall model fit to the data regarding current objectives (OSM S4). Details of posterior sampling procedures and diagnostics, full posterior samples and numerical summaries, and full model code, are in OSM S5.

Sex-, year- and tactic-specific switching and survival probabilities are directly estimated by the model. However, the relative population frequencies (i.e. proportions) of each migratory tactic in each year, and hence the degree of among-individual variation in flexibility, are not directly estimated. We therefore calculated the posterior distributions of tactic frequencies at the start of each year $y$ once new individuals have entered the dataset ($F_{1,y}$). Since $F_{1,y}$ jointly results from tactics initially undertaken by new individuals and from tactic switching of surviving individuals, it can be derived from the numbers of individuals that entered the dataset up to and including the current year and corresponding initial tactic, survival and tactic switching probabilities (OSM S3). Moreover, to synthesise how overall phenotypic changes from year $y$ to $y+1$ result from sequential within-year processes, we also calculated tactic frequencies immediately after annual survival selection ($F_{2,y}$), and after subsequent tactic switching ($F_{3,y}$; fig. 7A; OSM S3). We then derived the net effects on relative tactic frequencies of survival selection ($E_{1,y}=F_{2,y}-F_{1,y}$), supraflexibility ($E_{2,y}=F_{3,y}-F_{2,y}$), and entry of new individuals into the adult population ($E_{3,y}=F_{1,y+1}-F_{3,y}$; fig. 7A).

To further summarize year-dependent parameters, we also derived grand means and standard deviations across years (OSM S3). Estimates from the first study year

(2009-10) are not shown, and were not included in cross-year summaries, because low resighting probabilities prevented reliable inference on individual tactics (OSM S2). However, 2009-10 was retained in the dataset to initialise the model and differentiate subsequent entry of new individuals from individuals already present in the data (OSM S2). We present all estimates of directly estimated and derived parameters as posterior means with 95% credible intervals ('95%CI').

To explicitly test for between-tactic, between-sex and between-year differences in tactic frequencies, switching probabilities and survival probabilities, we computed the posterior distributions of such differences (Δ). We assessed support for the sign of each difference through the posterior probability that it was positive ($\Pr(\Delta>0)$; values close to 1 or 0 indicate strong evidence for positive or negative differences respectively, whilst values close to 0.5 indicate no clear evidence for either). Finally, for each year and sex, we assessed support for the four possible shapes of survival selection on the annual migratory tactic along the R–M continuum (fig. 2B). More precisely, we calculated the posterior probability that selection was stabilising ($\Pr(\Phi_{full\text{-}R}<\Phi_{mixed\text{-}RM}>\Phi_{full\text{-}M})$), disruptive ($\Pr(\Phi_{full\text{-}R}>\Phi_{mixed\text{-}RM}<\Phi_{full\text{-}M})$), directional towards full-R ($\Pr(\Phi_{full\text{-}R}>\Phi_{mixed\text{-}RM}>\Phi_{full\text{-}M})$), or directional towards full-M ($\Pr(\Phi_{full\text{-}R}<\Phi_{mixed\text{-}RM}<\Phi_{full\text{-}M})$). These four probabilities sum to 1, and particular probabilities close to 1 indicate strong evidence for the corresponding shape of selection.

# Results

## *Tactic frequencies and flexibility within years*

All three defined migratory tactics, and hence flexible and inflexible phenotypes, occurred in substantial proportions among individuals alive at the start of each focal biological year (fig. 4). Full-R was typically the most frequent tactic in both sexes (grand mean proportions across years with 95%CI: 0.42 [0.38,0.46] in females, 0.43 [0.39,0.47] in males; fig. 4; see OSM S5 for details). Meanwhile, mixed-RM was slightly less frequent than full-M in females (grand mean proportions: 0.26 [0.23,0.30] and 0.31 [0.28,0.34], respectively), but more frequent in males (0.32 [0.28,0.36] and 0.25 [0.22,0.28], respectively; fig. 4; OSM S5).

There was also substantial among-year variation in tactic frequencies, particularly for full-R (fig. 4; OSM S5). Interestingly, the two years where phenotypic effects of ECEs could potentially be detected (2012-13 and 2013-14) showed relatively low frequencies of full-R, with no evidence that full-R was more frequent than the other two tactics (fig. 4; OSM S5). Together, such within-year and among-year variation indicate substantial opportunity for selection and switching among flexible and inflexible tactics to drive appreciable shifts in phenotype frequencies, including effects of ECEs.

While there was no clear evidence that the frequency of full-R differed between males and females in any year, there was strong evidence for lower frequency of mixed-RM and/or higher frequency of full-M in females than males in several years, and hence on average across years (fig. 4; OSM S5). This indicates slight sexual dimorphism in migratory tactic, implying slight dimorphism in within-winter phenotypic flexibility.

### *Supraflexibility between years*

Probabilities of switching tactics between consecutive years revealed structured patterns of supraflexibility. Overall, probabilities of repeating the same tactic (i.e. not switching to another tactic) were notably high (fig. 5). These probabilities were highest for full-R (grand mean: 0.86 [0.81,0.90] and 0.87 [0.83,0.90] in females and males respectively), lowest for mixed-RM (grand mean: 0.54 [0.46,0.62] and 0.66 [0.60,0.73]), and intermediate for full-M (grand mean: 0.72 [0.68,0.76] and 0.74 [0.69,0.78]). There was strong evidence for this ranking of repeatabilities within most years, and hence in the grand mean across years, in both sexes (OSM S5). Individuals' migratory tactics were therefore notably highly repeatable, and the most flexible tactic was the least repeatable (hence, most supraflexible), as expected given the threshold trait model (fig. 2A).

Further, tactic switching was clearly structured along the R–M continuum. Specifically, full-R and full-M individuals switched to mixed-RM more than to full-M and full-R respectively (fig. 5; OSM S5). Grand mean switching probabilities were 0.10 [0.07,0.14] and 0.11 [0.08,0.14] from full-R to mixed-RM, and 0.24 [0.19,0.28] and 0.22 [0.17,0.26] from full-M to mixed-RM, in females and males respectively. In contrast, these probabilities were 0.04 [0.02,0.06] and 0.02 [0.01,0.04] from full-R to full-M, and 0.04 [0.02,0.07] and 0.05 [0.2,0.07] from full-M to full-R. Meanwhile, mixed-RM individuals switched in higher and roughly similar proportions to both inflexible tactics, with grand mean probabilities of 0.18 [0.13,0.24] and 0.19 [0.14,0.23] for switching to full-R, and 0.28 [0.22,0.34] and 0.15 [0.11,0.19] for switching to full-M, in females and males respectively (fig. 5; OSM S5). These relative probabilities of different forms of supraflexibility again qualitatively match expectations given the threshold trait model (fig. 2B).

Probabilities of tactic switching also varied among years (fig. 5; OSM S5). Interestingly, full-R females and males were especially likely to switch to another tactic in the two consecutive ECE years for which our model could detect coincident phenotypic changes (2012-13 and 2013-14, fig. 5). However, there was no clear evidence of a difference between these years, or with several non-ECE years (2011-12 and 2017-18 in females, 2014-15 and 2018-19 in males; fig. 5; OSM S5). Large among-year variation in switching probabilities also occurred for other tactics. Notably, mixed-RM showed strong evidence for changes in the magnitude and direction of switching in both sexes (OSM S5).

Finally, in several years, there was some or strong evidence for differences in switching probabilities between males and females (fig. 5, OSM S5). For example, females were often more likely to switch from mixed-RM to full-M than males, while full-M males showed noticeably consistent switching probabilities compared to females (fig. 5, OSM S5). Supraflexibility was therefore somewhat sexually dimorphic.

### *Survival selection on tactics and flexibility*

Our analyses revealed notable variation in the shape and strength of survival selection on the defined migratory tactics, and hence on within-year flexibility, between the pooled non-ECE years and each ECE year, and between females and males (fig. 6).

In non-ECE years, survival probabilities were very high in both sexes (fig. 6). Nonetheless, in males, there was strong evidence for disruptive selection along the R–M continuum, such that full-R and full-M had higher survival probabilities than mixed-RM (Δ: 0.06 [0.00,0.13] and 0.08 [0.01,0.15], respectively; fig. 6). In contrast, there was no evidence that survival differed among tactics in females (Δ: 0.01 [-0.06,0.08] and 0.03 [-0.05,0.12]; fig. 6).

In 2012-13, the first ECE year, there was strong evidence for directional selection towards full-M along the R–M continuum in males, and considerable evidence for a similar pattern in females (fig. 6). In both sexes, survival probability was higher for full-M than mixed-RM (Δ: 0.14 [-0.01,0.29] in females, 0.16 [0.00,0.32] in males; fig. 6), and even higher than for full-R (Δ: 0.20 [0.08,0.32] in females, 0.33 [0.21,0.46] in males). However, survival probability was lower for full-R than mixed-RM in males (Δ: -0.17 [-0.31,-0.03]), but this was not evident in females (Δ: -0.06 [-0.22,0.10]; fig. 6).

In 2013-14, there was strong evidence for disruptive selection in females (fig. 6), such that both full-R and full-M had higher survival probability than mixed-RM (Δ: 0.31 [0.09,0.52] and 0.18 [-0.02,0.38], respectively; fig. 6). There was also some evidence of additional skew in the shape of selection, with higher survival probability in full-R than full-M (Δ: 0.13 [-0.04,0.29]; fig. 6). In contrast, in males there was more evidence for stabilising selection than for the other possible shapes of selection, but no strong evidence that survival probability differed between tactics (fig. 6).

In 2017-2018, there was some evidence for directional selection towards full-M in females, and for stabilising selection in males (fig. 6). More precisely, in both sexes, there was some or strong evidence that full-R had lower survival probability than mixed-RM (Δ: -0.14 [-0.30,0.04] in females, -0.21 [-0.33,-0.06] in males; fig. 6) and full-M (Δ: -0.22 [-0.34,-0.10] in females, -0.13 [-0.25,-0.02] in males). However, there was no clear evidence for differences in survival probability between full-M and mixed-RM (Δ: 0.08 [-0.10,0.29] in females, -0.07 [-0.22,0.11] in males; fig. 6).

### *Net effects on phenotypic dynamics*

The three sequential processes of selection, supraflexibility and entry of new individuals each caused detectable immediate changes in tactic frequencies, and hence in the degree of within-winter migratory flexibility, across multiple years in each sex (fig. 7). On average across all years, supraflexibility caused the largest changes in tactic frequencies, while survival selection and entry of individuals each caused changes of smaller but similar sizes (fig. 7; OSM S5). Yet, these changes varied considerably in size and direction among years (fig. 7; OSM S5). Within-generation effects of survival selection on phenotype frequencies were consequently either accentuated or counter-acted (or even cancelled) by subsequent supraflexibility in different instances. For example, the reduced frequency of full-R associated with the 2012-13 ECE resulted from both supraflexibility (between year 2011-12 and 2012-13) and survival selection (through 2012-13), but was quickly reversed across subsequent years (fig. 7; OSM S5).

## Discussion

Within-individual and among-individual variation in phenotype, and associated variation in fitness components, underpins joint phenotypic, population dynamic and micro-evolutionary responses to environmental variation and change (Nussey et al. 2008; Chevin et al. 2010; Coulson et al. 2010). Yet, hierarchical patterns of individual phenotypic flexibility within and across biologically-relevant time periods, and associated episodes of directional, stabilising or disruptive selection affecting flexibility, are rarely quantified. Our analyses of ten years of large-scale band-resighting data revealed substantial among-individual variation and structure in flexibility and supraflexibility of seasonal migration versus residence, a key trait that directly shapes spatio-seasonal population dynamics, in European shags. We also demonstrate forms of disruptive and directional survival selection on migratory tactics along the residence-migration continuum, that varied among years and occasionally between the sexes. We thus illustrate that hierarchical variation in flexibility across timescales in dichotomous phenotypes can be structured in accordance with principles of threshold trait expression and can be subject to complex and varying selection, thereby shaping population-wide phenotypic dynamics.

### *Hierarchical variation in flexibility*

Our focal adult shags showed substantial among-individual variation yet notably high within-individual repeatability in annual tactics, including high repeatability of both seasonally inflexible phenotypes (full-R and full-M) and seasonally flexible phenotypes (mixed-RM, fig. 4, 5). This implies that both inflexibility and flexibility in the focal threshold trait could be underpinned by substantial genetic and/or permanent environmental effects.

To our knowledge, four previous studies have quantified within-individual variation in plasticity across different time periods and/or environmental axes alongside among-individual variation (Araya-Ajoy and Dingemanse 2017; Mitchell and Biro 2017; Cornwell et al. 2019; Strickland et al. 2021). All four studies considered continuously distributed behavioural traits and quantified phenotypic variation as reaction norms slopes given known environmental variables, finding considerable repeatability through time. In general, such repeatability could readily occur if plastic responses are repeated because forms of environmental variation that drive reaction norms are repeated through time, for example due to environmental cyclicity such as seasonality (Schlichting and Pigliucci 1998; Piersma and Drent 2003; Lande et al. 2019). Any underlying permanent environmental effects might reflect habituation or social learning for behaviours including migration (Chapman et al. 2011; Jesmer et al. 2018). Repeatability in phenotypic plasticity could also occur in non-repeated environmental contexts if individual reaction norm slopes are consistent across different environmental axes (e.g. Cornwell et al., 2019), as expected if common physiological or cognitive machineries underpin plasticity in multiple contexts (Coppens et al. 2010; Dingemanse and Wolf 2013; Stamps 2016). Accordingly, within-individual phenotypic variation has been interpreted to result from responses to internal or external variables that vary through time at multiple scales within a population (Araya-Ajoy and Dingemanse 2017; Mitchell and Biro 2017; Cornwell et al. 2019; Strickland et al. 2021), representing multidimensional plasticity (Westneat et al. 2015, 2019).

In contrast, the threshold trait model, where an underlying environmentally-sensitive liability leads to discrete phenotypic outcomes (fig. 1), can directly predict structured patterns of hierarchical variation and repeatability in flexibility across timescales without necessarily requiring multidimensional plasticity. Specifically, because the degree of individual phenotypic flexibility is expected to decrease with increasing distance of initial liability from the threshold (i.e. with liability elevation, Reid and Acker 2022), among-individual variation in flexibility with directional supraflexibility and positive covariation between flexibility and supraflexibility (fig. 1, 2) can readily arise, as observed for migration versus residence in shags (fig. 4, 5).

Some studies on migration versus residence in other taxa also provide evidence of among-individual variation in within-year flexibility, but did not explicitly consider within-individual repeatability of such flexibility (e.g. roe deer, *Capreolus capreolus*: Cagnacci et al. 2011; blackbird, *Turdus merula*: Fudickar et al. 2013; red deer, *Cervus elaphus*: Peters et al. 2019). Other studies demonstrate high repeatability of migration versus residence with some between-year flexibility, but did not explicitly consider within-year flexibility (e.g. American dipper, *Cinclus mexicanus*: Gillis et al. 2008; red-spotted newt: *Notophtalmus viridescens*: Grayson et al. 2011; skylark, *Alauda arvensis*: Hegemann et al. 2015, noctule bat *Nyctalus noctule*: Lehnert et al. 2018; various ungulates: Berg et al. 2019). Following our current analyses, future studies on these and other systems could explicitly quantify structured patterns of hierarchical variation in phenotypic flexibility, and hence assess general accordance with the threshold trait model. Environmental variables known to be associated with cross-sectional changes in migratory proportions (e.g. individual condition, conspecific density, climate variables, or resource availability; Brodersen et al. 2008; Skov et al. 2010; Grayson and Wilbur 2009; Boyle et al. 2010; Eggeman et al. 2016) would then be relevant candidates to explicitly consider sex-specific liability-scale reaction norms underlying phenotypic flexibility and supraflexibility and hence formally link concepts of observed-scale flexibility and latent liability-scale plasticity.

### *Selection on flexibility*

Our analyses also demonstrate occurrences of notably strong survival selection on annual migratory tactics and hence on migratory flexibility, with substantial variation in the shape and strength of selection between non-extreme years and each ECE year, and sometimes between females and males (fig. 6). Selection against full-R during the ECE years 2012-13 and 2017-18 reflects greater exposure of resident versus migrant locations to extreme late-winter storms that caused mortality in these two years (broadly consistent with previous cross-sectional analyses that did not consider annual migratory tactics; Acker et al. 2021a). The lower survival probability of mixed-RM compared to full-M in 2012-13, generating overall directional selection towards full-M, could then effectively represent a cost of flexibility (here, of moving between locations during winter) that prevented the full potential benefit of migration. The disruptive selection along the R–M continuum that was evident for the five non-extreme years in males, and in one ECE year in females (2013-14; fig. 6), is also consistent with costs of mixed-RM. In general, such costs might result from fine spatio-temporal variation in survival if mixed-RM individuals experience temporarily poor conditions both before and after departure. They could also represent direct costs resulting from reduced local site familiarity or increased susceptibility to local environmental deteriorations compared with individuals that winter in a single location (Skov et al. 2010), and/or from maintaining physiological machineries required to respond to environmental variation (e.g. DeWitt et al. 1997; Auld et al. 2010).

Still relatively few studies have quantified variation in fitness components in relation to phenotypic plasticity in any trait, and very few considered labile traits. Conclusive evidence of selection on phenotypic plasticity, or on resulting phenotypic flexibility, consequently remains scant (Van Buskirk and Steiner 2009; Chevin et al. 2010; Arnold et al. 2019). A general meta-analysis concluded that fitness costs and benefits of developmental plasticity were usually small, but that costs predominantly arise under stressful environmental conditions (Van Buskirk and Steiner 2009). In contrast, our results indicate that costs of flexibility (i.e. lower survival of mixed-RM, resulting in disruptive selection) can be manifested under both benign (non-ECE) and extreme (ECE) conditions (fig. 6). Yet, our results also show that any such costs are highly contingent, and may often be hidden by large variation in sex- and year-dependent costs and benefits of the alternative inflexible phenotypic tactics (i.e. full-M and full-R).

Conversely, there was some evidence of stabilising selection, defined by higher survival probability of the flexible mixed-RM tactic than both inflexible tactics, only in 2013-14 and 2017-18 in males. There was consequently no strong evidence that the relatively high observed frequency of migratory flexibility could be directly maintained through survival selection. However, the threshold trait model (fig. 1) implies that flexibility could be induced by instances of phenotypically disassortative mating. This can arise because any reproduction involving two parents with opposite inflexible phenotypes, given additive genetics effects on their liabilities, can readily generate offspring with liabilities close to the threshold that are consequently highly phenotypically flexible (Reid and Acker 2022). Variation in phenotypic flexibility in a threshold trait might then be maintained as an indirect consequence of disruptive selection that acts to maintain both inflexible phenotypes.

Indeed, under the threshold trait model, any form of selection along the continuum of observed phenotypic tactics implies some similar form of selection along the underlying liability (fig. 1,2; Reid and Acker 2022). While substantial additive genetic variation in liabilities can be maintained even given consistent directional selection, phenotypic variation is unlikely to be maintained without some additional form of non-linear or varying selection (Roff 1996, 1998). These requirements could be effectively fulfilled by the disruptive selection observed in non-ECE years, and the variation in the shape of selection observed across ECE years, in turn implying that increasing frequencies of diverse ECEs could help maintain genetic variation in flexibility. Yet, the frequency of full-R typically exceeded other phenotypic tactics, despite having no clear survival advantage across years. This may partly be explained by reproductive selection, since our previous cross-sectional analyses showed that shags that were resident on Isle of May during late winter commonly had higher breeding success than migrants that spent the late winter elsewhere (Grist et al. 2017; Acker et al. 2021b). Such relationships could represent carry-over effects of winter location on subsequent reproductive success, generating indirect selection on migratory tactic.

Meanwhile, few or no previous studies on any trait have considered whether selection on phenotypic plasticity or flexibility is sex-specific (Van Buskirk and Steiner 2009; Arnold et al. 2019; but see Fox et al. 2019). We observed a striking difference in the form of selection between females and males in the ECE year 2013-14, with apparently strong selection against migratory flexibility in females but not males. There was also some weak evidence of selection for flexibility in males but not females in 2017-18. These effects could reflect interactions between environmental variation and sex-specific environmental sensitivity in shags. For example, females' foraging efficiency is more sensitive to strong winds, and females can experience higher associated mortality (Lewis et al. 2015; Acker et al. 2021a). Given the threshold trait model, such sex-specific selection on flexibility could translate into sex-specific selection on mean liability, due to the expected intrinsic relationship between the two. Thus, while sex-specific selection on any continuously distributed trait could cause or accentuate sexual dimorphism (Kruuk et al. 2008; Cox and Calsbeek 2009), such effects could be exacerbated in a threshold trait if sex-specific selection affects flexibility. However, since there was no strong sexual dimorphism in the shape of selection in most years, any such evolutionary consequences are likely to be minor in our system, unless ECEs become more frequent.

### *Overall implications for phenotypic dynamics*

One general yet rarely achieved ambition in evolutionary ecology is to partition effects of multiple simultaneously or sequentially acting processes on overall phenotypic variation and change (e.g. Charmantier et al. 2008; Chevin et al. 2010; Coulson et al. 2010; de Villemereuil et al. 2020). Our analyses allowed evaluation of the relative contributions of supraflexibility, survival selection on flexibility and entry of new adult individuals in shaping population-wide phenotypic change among adults across the study period, including change in flexibility. Despite the strong observed episodes of selection, our results demonstrate some predominance of supraflexibility over selection on flexibility in generating immediate changes in tactic frequencies, and hence in shaping population-wide occurrences of flexible seasonal migration and resulting spatio-seasonal dynamics (fig. 7). Yet, selection could still shape longer-term micro-evolutionary and phenotypic dynamics by altering within-generation distributions of genetic values for liability-scale reaction norm elevations and/or slopes and corresponding frequencies of flexible and inflexible tactics, and of associated supraflexibility, in subsequent generations. Such selection could partly account for the observed effects of entry of new adults, which are primarily new recruits to the local breeding population, and hence represent contributions from successive (overlapping) generations.

The full longer-term dynamics of phenotypes, and associated degrees of flexibility and supraflexibility, will consequently depend on additive genetic variances in liability for migration versus residence and patterns of (dis)assortative mating, alongside developmental plasticity and selection occurring through reproduction and sub-adult survival to recruitment (Charlesworth 1994; Kruuk 2004; Coulson et al. 2010; Reid and Acker 2022). These dynamics could be further modulated by complex forms of direct and indirect selection on supraflexibility, including indirect selection resulting from the expected positive covariation with flexibility (fig. 2C). Our accumulating dataset will allow future estimation of all these components, but this will require further major methodological developments. New multistate capture-recapture model structures will be required to appropriately represent memory of the previous phenotypic tactic, reproduction, early-life histories up to recruitment, and phenotype-dependent mating dynamics. Combining capture-recapture and generalized linear mixed model approaches will ultimately allow estimation of effects on the latent liability scale, including plasticity reaction norms and additive genetic variances. Together, these advances will allow full dissection of all processes shaping short-term and longer-term changes in migratory phenotypes and migratory flexibility, and resulting spatio-seasonal population dynamics.

## Acknowledgements

We thank everyone from UK Centre for Ecology & Hydrology and University of Aberdeen (UoA) who contributed to data collection, particularly Hannah Grist and Jenny Sturgeon; NatureScot for access to the Isle of May National Nature Reserve; Scottish Ornithologists' Club (SOC) for their support; and all volunteer observers, particularly Raymond Duncan and Moray Souter. This work was funded by Natural Environment Research Council (NERC; award NE/R000859/1 and award

NE/R016429/1 as part of the UK-SCaPE programme delivering National Capability), Norwegian Research Council (SFF-III grant 223257), NTNU, UK Royal Society, Marine Alliance for Science and Technology for Scotland (MASTS), SOC and UoA.

## Statement of authorship

PA and JMR formulated the ideas and conceptual developments. JMR and FD conceived the migration field study. FD, SW, MPH, MAN, SJB, CG, RS and JMR organised and undertook long-term field data collection. PA designed the modelling methodology, coded the models, analysed the data and drafted the manuscript, assisted by JMR. FD, SJB, and APP contributed manuscript edits. All authors gave final approval for publication.

## Data and code accessibility

Data and code are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.3j9kd51mn.

## Literature cited

Acker, P., S. J. Burthe, M. A. Newell, H. Grist, C. Gunn, M. P. Harris, A. Payo-Payo, et al. 2021*a*. Episodes of opposing survival and reproductive selection cause strong fluctuating selection on seasonal migration versus residence. Proceedings of the Royal Society B 288:20210404.

Acker, P., F. Daunt, S. Wanless, S. J. Burthe, M. A. Newell, M. P. Harris, H. Grist, et al. 2021*b*. Strong survival selection on seasonal migration versus residence induced by extreme climatic events. Journal of Animal Ecology 90:796–808.

Aebischer, N. J. 1995. Philopatry and colony fidelity of Shags *Phalacrocorax aristotelis* on the east coast of Britain. Ibis 137:11–18.

Araya-Ajoy, Y. G., and N. J. Dingemanse. 2017. Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. Journal of Animal Ecology 86:227–238.

Araya-Ajoy, Y. G., K. J. Mathot, and N. J. Dingemanse. 2015. An approach to estimate short-term, long-term and reaction norm repeatability. Methods in Ecology and Evolution 6:1462–1473.

Arnold, P. A., A. B. Nicotra, and L. E. B. Kruuk. 2019. Sparse evidence for selection on phenotypic plasticity in response to temperature. Philosophical Transactions of the Royal Society B 374:20180185.

Arnold, S. J., and M. J. Wade. 1984. On the measurement of natural and sexual selection: theory. Evolution 38:709–719.

Auld, J. R., A. A. Agrawal, and R. A. Relyea. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proceedings of the Royal Society B 277:503–511.

Barlow, E. J., F. Daunt, S. Wanless, and J. M. Reid. 2013. Estimating dispersal distributions at multiple scales: within-colony and among-colony dispersal rates, distances and directions in European Shags *Phalacrocorax aristotelis*. Ibis 155:762–778.

Berg, J. E., M. Hebblewhite, C. C. St. Clair, and E. H. Merrill. 2019. Prevalence and mechanisms of partial migration in ungulates. Frontiers in Ecology and Evolution 7:325.

Berthold, P. 1988. Evolutionary aspects of migratory behavior in European warblers. Journal of Evolutionary Biology 1:195–209.

Boyle, W. A., D. R. Norris, and C. G. Guglielmo. 2010. Storms drive altitudinal migration in a tropical bird. Proceedings of the Royal Society B 277:2511–2519.

Brodersen, J., P. A. Nilsson, L.-A. Hansson, C. Skov, and C. Brönmark. 2008. Condition-dependent individual decision-making determines cyprinid partial migration. Ecology 89:1195–1200.

Brommer, J. E. 2013. Phenotypic plasticity of labile traits in the wild. Current Zoology 59:485–505.

Cagnacci, F., S. Focardi, M. Heurich, A. Stache, A. J. M. Hewison, N. Morellet, P. Kjellander, et al. 2011. Partial migration in roe deer: migratory and resident tactics are end points of a behavioural gradient determined by ecological factors. Oikos 120:1790–1802.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, et al. 2017. Stan: a probabilistic programming language. Journal of Statistical Software 76:1–32.

Chapman, B. B., C. Brönmark, J.-Å. Nilsson, and L.-A. Hansson. 2011. The ecology and evolution of partial migration. Oikos 120:1764–1775.

Charlesworth, B. 1994. Evolution in age-structured populations. 2nd ed. Cambridge University Press, Cambridge.

Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk, and B. C. Sheldon. 2008. Adaptive phenotypic plasticity in response to climate change in a wild bird population. Science 320:800–803.

Chevin, L.-M., R. Lande, and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLOS Biology 8:e1000357.

Coppens, C. M., S. F. de Boer, and J. M. Koolhaas. 2010. Coping styles and behavioural flexibility: towards underlying mechanisms. Philosophical Transactions of the Royal Society B 365:4021–4028.

Cornwell, T. O., I. D. McCarthy, C. R. A. Snyder, and P. A. Biro. 2019. The influence of environmental gradients on individual behaviour: Individual plasticity is consistent across risk and temperature gradients. Journal of Animal Ecology 88:511–520.

Coulson, T., S. Tuljapurkar, and D. Z. Childs. 2010. Using evolutionary demography to link life history theory, quantitative genetics and population ecology. Journal of Animal Ecology 79:1226–1240.

Cox, R. M., and R. Calsbeek. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. The American Naturalist 173:176–187.

Daunt, F., T. E. Reed, M. Newell, S. Burthe, R. A. Phillips, S. Lewis, and S. Wanless. 2014. Longitudinal bio-logging reveals interplay between extrinsic and intrinsic carry-over effects in a long-lived vertebrate. Ecology 95:2077–2083.

DeWitt, T. J., A. Sih, and D. S. Wilson. 1998. Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution 13:77–81.

Dingemanse, N. J., and M. Wolf. 2013. Between-individual differences in behavioural plasticity within populations: causes and consequences. Animal Behaviour, 85:1031–1039.

Dingle, H. 1996. Migration: the biology of life on the move. Oxford University Press, Oxford.

Dodson, J. J., N. Aubin-Horth, V. Thériault, and D. J. Páez. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. Biological Reviews 88:602–625.

Eggeman, S. L., M. Hebblewhite, H. Bohm, J. Whittington, and E. H. Merrill. 2016. Behavioural flexibility in migratory behaviour in a long-lived large herbivore. Journal of Animal Ecology 85:785–797.

Falconer, D. S., and T. F. C. Mackay. 1996. Quantitative genetics. 4th ed. Prentice Hall, Harlow.

Fox, R. J., J. M. Donelson, C. Schunter, T. Ravasi, and J. D. Gaitán-Espitia. 2019. Beyond buying time: the role of plasticity in phenotypic adaptation to rapid environmental change. Philosophical Transactions of the Royal Society B 374:20180174.

Fudickar, A. M., A. Schmidt, M. Hau, M. Quetting, and J. Partecke. 2013. Female-biased obligate strategies in a partially migratory population. Journal of Animal Ecology 82:863–871.

Gabriel, W., B. Luttbeg, A. Sih, and R. Tollrian. 2005. Environmental tolerance, heterogeneity, and the evolution of reversible plastic responses. The American Naturalist 166:339–353.

Gelman, A., X.-L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica 6:733–760.

Gianola, D. 1982. Theory and analysis of threshold characters. Journal of Animal Science 54:1079–1096.

Gienapp, P., C. Teplitsky, J. S. Alho, J. A. Mills, and J. Merilä. 2008. Climate change and evolution: disentangling environmental and genetic responses. Molecular Ecology 17:167–178.

Gillis, E. A., D. J. Green, H. A. Middleton, and C. A. Morrissey. 2008. Life history correlates of alternative migratory strategies in American dippers. Ecology 89:1687–1695.

Gimenez, O., A. Viallefont, A. Charmantier, R. Pradel, E. Cam, C. R. Brown, M. D. Anderson, M. B. Brown, R. Covas, and J.-M. Gaillard. 2008. The risk of flawed inference in evolutionary studies when detectability is less than one. American Naturalist 172:441–448.

Grayson, K. L., L. L. Bailey, and H. M. Wilbur. 2011. Life history benefits of residency in a partially migrating pond-breeding amphibian. Ecology 92:1236–1246.

Grayson, K. L., and H. M. Wilbur. 2009. Sex- and context-dependent migration in a pond-breeding amphibian. Ecology 90:306–312.

Grist, H., F. Daunt, S. Wanless, S. J. Burthe, M. A. Newell, M. P. Harris, and J. M. Reid. 2017. Reproductive performance of resident and migrant males, females and pairs in a partially migratory bird. Journal of Animal Ecology 86:1010–1021.

Grist, H., F. Daunt, S. Wanless, E. J. Nelson, M. P. Harris, M. Newell, S. Burthe, et al. 2014. Site fidelity and individual variation in winter location in partially migratory European shags. PLOS ONE 9:e98562.

Hadfield, J. D. 2008. Estimating evolutionary parameters when viability selection is operating. Proceedings of the Royal Society B 275:723–734.

Hegemann, A., P. P. Marra, and B. I. Tieleman. 2015. Causes and consequences of partial migration in a passerine bird. The American Naturalist 186:531–546.

Hendry, A. P. 2016. Key questions on the role of phenotypic plasticity in eco-evolutionary dynamics. Journal of Heredity 107:25–41.

Jesmer, B. R., J. A. Merkle, J. R. Goheen, E. O. Aikens, J. L. Beck, A. B. Courtemanch, M. A. Hurley, et al. 2018. Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. Science 361:1023–1025.

Keogan, K., S. Lewis, R. J. Howells, M. A. Newell, M. P. Harris, S. Burthe, R. A. Phillips, et al. 2021. No evidence for fitness signatures consistent with increasing trophic mismatch over 30 years in a population of European shag *Phalacrocorax aristotelis*. Journal of Animal Ecology 90:432–446.

Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the ‘animal model’. Philosophical Transactions of the Royal Society B 359:873–890.

Lande, R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292-305.

Lande, R. 2019. Developmental integration and evolution of labile plasticity in a complex quantitative character in a multiperiodic environment. Proceedings of the National Academy of Sciences 116:11361–11369.

Lehnert, L. S., S. Kramer-Schadt, T. Teige, U. Hoffmeister, A. Popa-Lisseanu, F. Bontadina, M. Ciechanowski, et al. 2018. Variability and repeatability of noctule bat migration in Central Europe: evidence for partial and differential migration. Proceedings of the Royal Society B 285:20182174.

Lewis, S., R. A. Phillips, S. J. Burthe, S. Wanless, and F. Daunt. 2015. Contrasting responses of male and female foraging effort to year-round wind conditions. Journal of Animal Ecology 84:1490–1496.

Lundberg, P. 1988. The evolution of partial migration in Birds. Trends in Ecology & Evolution 3:172–175.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer associates, New York.

Mitchell, D. J., and P. A. Biro. 2017. Is behavioural plasticity consistent across different environmental gradients and through time? Proceedings of the Royal Society B 284:20170893.

Nakagawa, S., and R. P. Freckleton. 2008. Missing inaction: the dangers of ignoring missing data. Trends in Ecology & Evolution 23:592–596.

Newton, I. 2008. The migration ecology of birds. Elsevier, New York.

Nussey, D. H., A. J. Wilson, and J. E. Brommer. 2007. The evolutionary ecology of individual phenotypic plasticity in wild populations. Journal of Evolutionary Biology 20:831–844.

Piersma, T., and J. Drent. 2003. Phenotypic flexibility and the evolution of organismal design. Trends in Ecology & Evolution 18:228–233.

Pulido, F. 2011. Evolutionary genetics of partial migration – the threshold model of migration revis(it)ed. Oikos 120:1776–1783.

Pulido, F., P. Berthold, and A. J. van Noordwijk. 1996. Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. Proceedings of the National Academy of Sciences 93:14642–14647.

Reid, J. M., and P. Acker. 2022. Properties of phenotypic plasticity in discrete threshold traits. Evolution *in press*.

Reid, J. M., M. Souter, S. R. Fenn, P. Acker, A. Payo-Payo, S. J. Burthe, S. Wanless, et al. 2020. Among-individual and within-individual variation in seasonal migration covaries with subsequent reproductive success in a partially migratory bird. Proceedings of the Royal Society B 287:20200928.

Reid, J. M., J. M. J. Travis, F. Daunt, S. J. Burthe, S. Wanless, and C. Dytham. 2018. Population and evolutionary dynamics in spatially structured seasonally varying environments. Biological Reviews 93:1578–1603.

Roff, D. A. 1996. The evolution of threshold traits in animals. The Quarterly Review of Biology 71:3–35.

Roff, D. A. 1998. Evolution of threshold traits: the balance between directional selection, drift and mutation. Heredity 80:25–32.

Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer associates, New York, NY.

Sih, A., M. C. O. Ferrari, and D. J. Harris. 2011. Evolution and behavioural responses to human-induced rapid environmental change. Evolutionary Applications 4:367–387.

Skov, C., K. Aarestrup, H. Baktoft, J. Brodersen, C. Brönmark, L.-A. Hansson, E. E. Nielsen, et al. 2010. Influences of environmental cues, migration history, and habitat familiarity on partial migration. Behavioral Ecology 21:1140–1146.

Snell-Rood, E. C., M. E. Kobiela, K. L. Sikkink, and A. M. Shephard. 2018. Mechanisms of plastic rescue in novel environments. Annual Review of Ecology, Evolution, and Systematics 49:331–354.

Stamps, J. A. 2016. Individual differences in behavioural plasticities. Biological Reviews 91:534–567.

Strickland, K., D. J. Mitchell, C. Delmé, and C. H. Frère. 2021. Repeatability and heritability of social reaction norms in a wild agamid lizard. Evolution 75:1953–1965.

Van Buskirk, J., and U. K. Steiner. 2009. The fitness costs of developmental canalization and plasticity. Journal of Evolutionary Biology 22:852–860.

de Villemereuil, P., A. Charmantier, D. Arlt, P. Bize, P. Brekke, L. Brouwer, A. Cockburn, et al. 2020. Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. Proceedings of the National Academy of Sciences 117:31969–31978.

Westneat, D. F., J. Wright, and N. J. Dingemanse. 2015. The biology hidden inside residual within-individual phenotypic variation. Biological Reviews 90:729–743.

Westneat, David. F., L. J. Potts, K. L. Sasser, and J. D. Shaffer. 2019. Causes and consequences of phenotypic plasticity in complex environments. Trends in Ecology & Evolution 34:555–568.

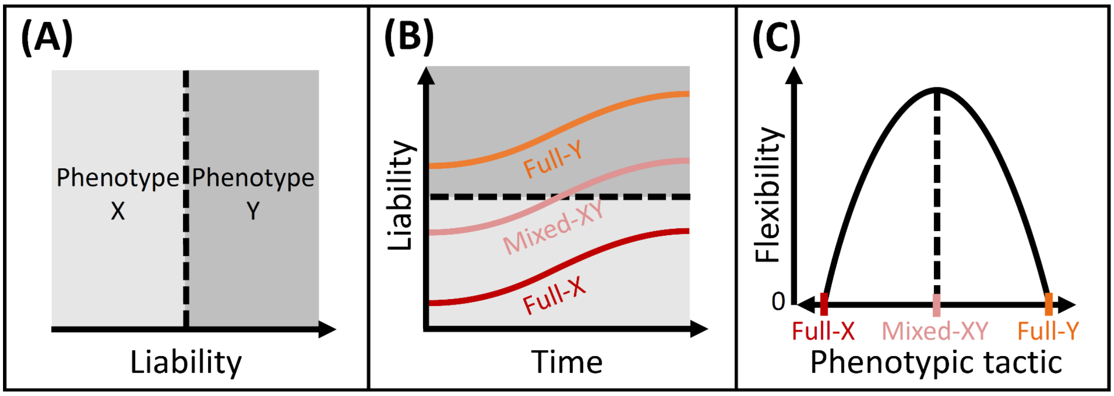


**Figure 1**. Basic principles of the threshold trait model. (A) Each individual expresses alternative phenotype X or Y (light grey versus dark grey shading) when its underlying liability is below or above the threshold (dashed line). Liability is conceptualised as a continuous latent variable which can vary among and within individuals due to genetic and environmental effects. (B) Hypothetical changes in individual liability through time, given among-individual variation in mean liability (i.e. elevation) and identical time-varying environmental effects. The red and orange individuals have liabilities that remain below and above the threshold (dashed line) respectively, and hence inflexibly express phenotypes X and Y respectively throughout the observed time period (i.e. full-X and full-Y tactics). The rose individual has a liability that crosses the threshold at some time, and hence expresses both X and Y (i.e. mixed-XY tactic, representing phenotypic flexibility). (C) Relationship between the degree of individual flexibility (i.e. within-individual phenotypic variance) and phenotypic tactic along the X–Y continuum (i.e. expressed proportion of X vs. Y, from full-X to full-Y). Flexibility is zero at both extremes and maximal for a mixed-XY tactic with 50% X and 50% Y (as illustrated).

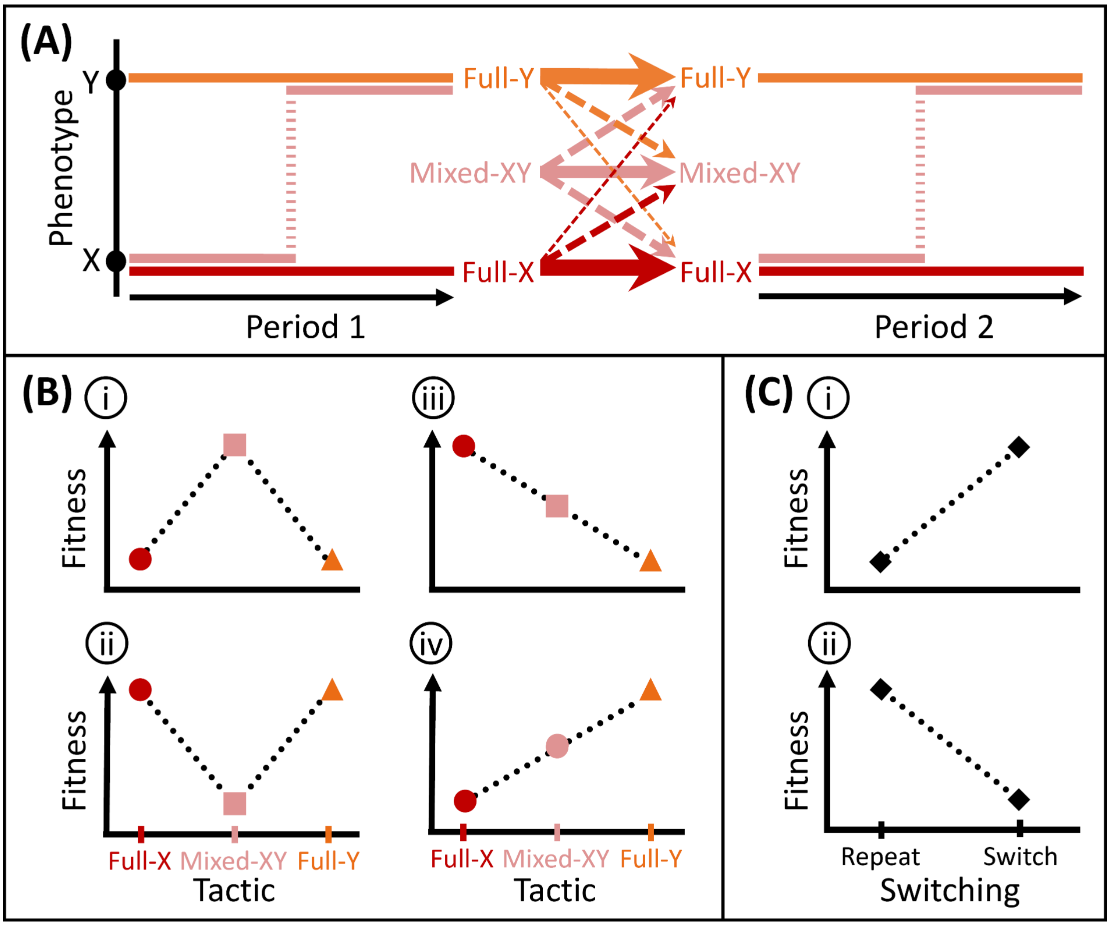


**Figure 2**. Hierarchical phenotypic variation in a dichotomous trait and associated selection. (A) Among-individual and within-individual variation in phenotypic tactic, illustrating variation in flexibility and supraflexibility. Phenotypes X and/or Y are expressed during focal time periods (denoted 1 and 2) that occur sequentially. Within each period, each individual expresses either an inflexible phenotypic tactic (full-X in red, or full-Y in orange) or a flexible tactic (mixed-XY in rose). There are nine possible paths across two consecutive periods and hence forms of supraflexibility (arrows; OSM S1): individuals can either retain the same tactic (3 possible paths, solid arrows), or switch tactic (6 possible paths, dashed arrows). Different arrow widths represent different switching probabilities as qualitatively expected under the threshold trait model, showing that supraflexibility could be structured and covary with flexibility. (B and C) Possible shapes of selection on tactics (and hence flexibility) and supraflexibility

arising due to associated variation in fitness ($\omega$), resulting in selection along the X–Y continuum. Such selection could be (B-i) stabilizing ($\omega_{\text{full-X}}<\omega_{\text{mixed-XY}}>\omega_{\text{full-Y}}$) or (B-ii) disruptive ($\omega_{\text{full-X}}>\omega_{\text{mixed-XY}}<\omega_{\text{full-Y}}$) due to relative benefits or costs of being flexible, or directional towards (B-iii) full-X ($\omega_{\text{full-X}}>\omega_{\text{mixed-XY}}>\omega_{\text{full-Y}}$) or (B-iv) full-Y ($\omega_{\text{full-X}}<\omega_{\text{mixed-XY}}<\omega_{\text{full-Y}}$) due to relative benefits of expressing X or Y. Further, there could be selection (C-i) for or (C-ii) against supraflexibility depending on fitness variation occurring in two consecutive periods. Such selection on supraflexibility could intrinsically result from sequences of selection on flexibility in consecutive years, and/or from direct costs or benefits of switching tactic between years. It could therefore be partly conditional on, but also partly independent of, the tactics from and to which switching occurs.

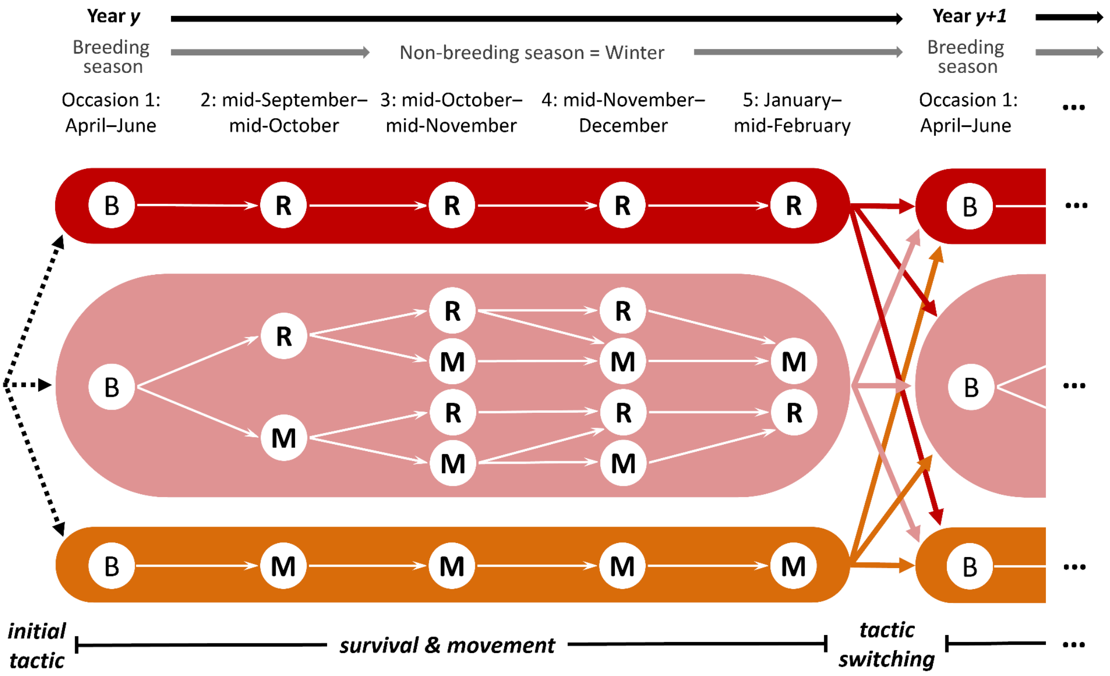


**Figure 3**. Summary of the state transition process of the full-annual-cycle capture-recapture model representing a dynamic finite mixture of annual migratory tactics, and hence individual variation in flexibility and supraflexibility of migration versus residence. The diagram represents possible fates of individuals from the year they enter the dataset (*y*) to the next year (*y+1*). An individual undertakes one possible initial tactic (dashed arrows): either full residence ('full-R', top, red), mixed residence-migration ('mixed-RM', middle, rose), or full migration ('full-M', bottom, orange). All individuals start the annual cycle during the breeding season at the residency area (state 'B'). Each tactic is defined by constraints that determine possible within-year transitions (white arrows) of surviving individuals between the two phenotypic states: residence (R) and migration (M). Then, between years (here year *y* to *y+1*), individuals can switch between tactics or repeat the same tactic (coloured arrows). The complete fate diagram showing the full state transition and observation processes is in OSM S3.

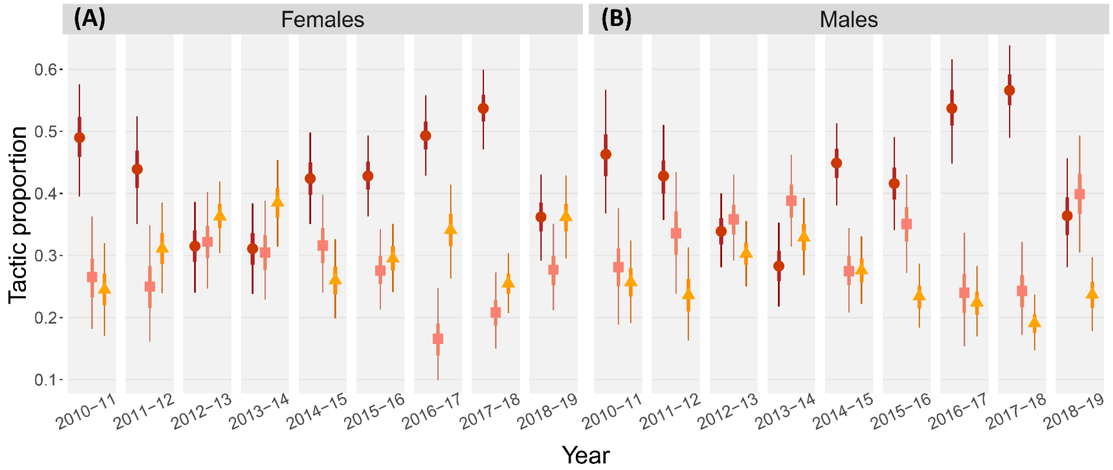


**Figure 4**. Relative tactic frequencies estimated at the start of each biological year (i.e. occasion 1; $F_1$) from 2010-11 to 2018-19 in (A) females and (B) males, derived for full residence (‘full-R’, red circles) and full migration (‘full-M’, orange triangles), representing tactics that are phenotypically inflexible within the focal winter, and mixed residence-migration (‘mixed-RM’, rose squares), representing a phenotypically flexible tactic within the focal winter. Point estimates are posterior means, inner and outer line segments indicate 50% and 95% credible intervals.

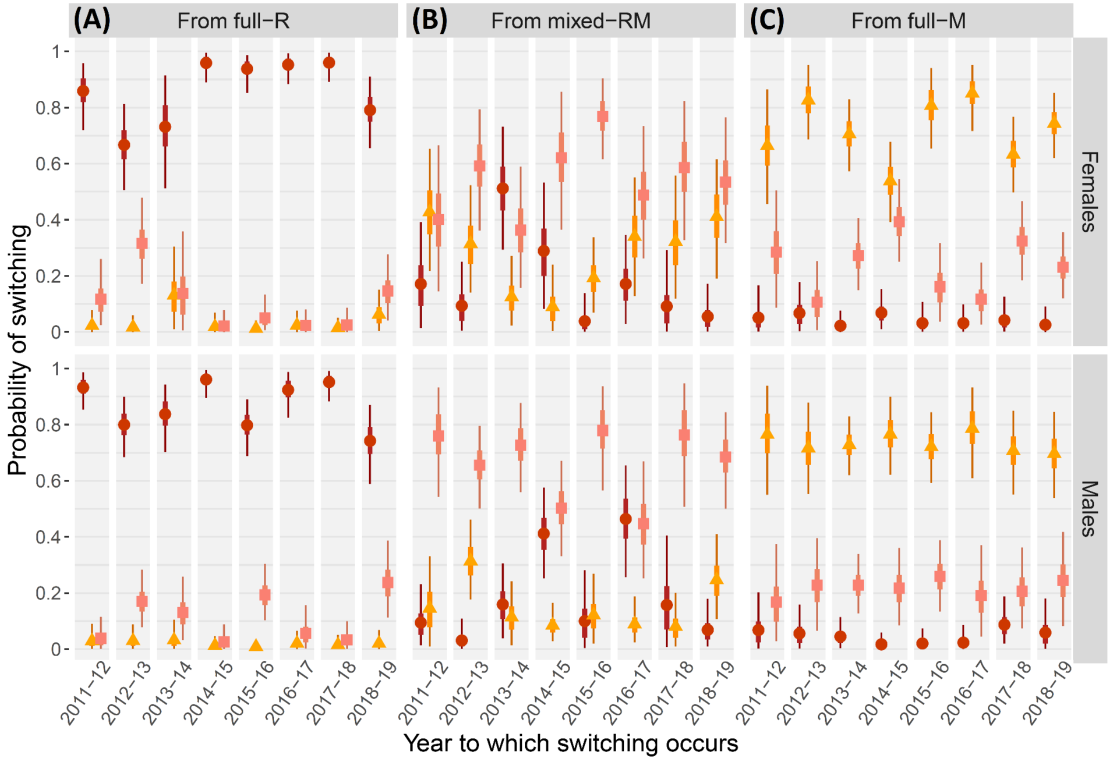
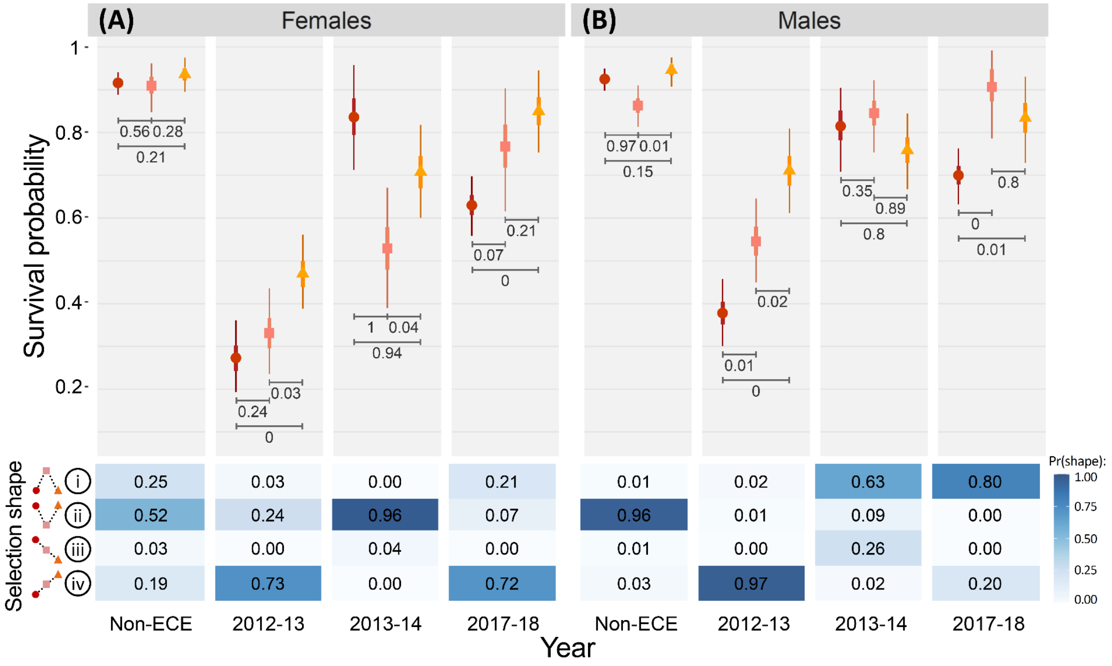


**Figure 5**. Probabilities of between-year switching ($\kappa$) from (A) full residence ('full-R'), (B) mixed residence-migration ('mixed-RM') and (C) full migration ('full-M') to each of these tactics, in females (top row) and males (bottom row). Probabilities of switching to full-R, mixed-RM and full-M are in red circles, rose squares and orange triangles respectively. The probability of switching from any one tactic to the same tactic represents the probability of repeating this tactic (i.e. not switching). Point estimates are posterior means, inner and outer line segments indicate 50% and 95% credible intervals. The x-axis indicates the biological year to which switching occurs (e.g. 2012-13 is for the transition from 2011-12 to 2012-13).



| Selection shape | (A) Females Non-ECE | (A) Females 2012-13 | (A) Females 2013-14 | (A) Females 2017-18 | (B) Males Non-ECE | (B) Males 2012-13 | (B) Males 2013-14 | (B) Males 2017-18 |
|---|---|---|---|---|---|---|---|---|
| ⓘ | 0.25 | 0.03 | 0.00 | 0.21 | 0.01 | 0.02 | 0.63 | 0.80 |
| ⓘⓘ | 0.52 | 0.24 | 0.96 | 0.07 | 0.96 | 0.01 | 0.09 | 0.00 |
| ⓘⓘⓘ | 0.03 | 0.00 | 0.04 | 0.00 | 0.01 | 0.00 | 0.26 | 0.00 |
| ⓘⓥ | 0.19 | 0.73 | 0.00 | 0.72 | 0.03 | 0.97 | 0.02 | 0.20 |

**Figure 6**. Tactic-specific annual survival probabilities, and associated evidence for possible shapes of selection, in (A) females and (B) males. Annual survival probability ($\Phi$) was estimated separately for each of the three years with an extreme climatic event ('ECE'), and estimated across all five non-ECE years combined. In the survival probability graphs, point estimates are posterior means for full residents ('full-R', red circles), mixed resident-migrants ('mixed-RM', rose squares), and full migrants ('full-M', orange triangles). Inner and outer line segments indicate 50% and 95% credible intervals. Segment annotations indicate the probability that the difference $\Delta$, between the two points at the left and right end of the segment, is positive (Pr($\Delta$>0); values close to 0 or 1 indicate substantial evidence for a negative or positive difference respectively; values close to 0.5 indicate no clear evidence for either). In the heatmaps, values are the probabilities of each possible shape of selection along the residence–migration continuum (Pr(shape); i: stabilising, ii: disruptive, iii: directional towards full-R, and iv: directional towards full-M).

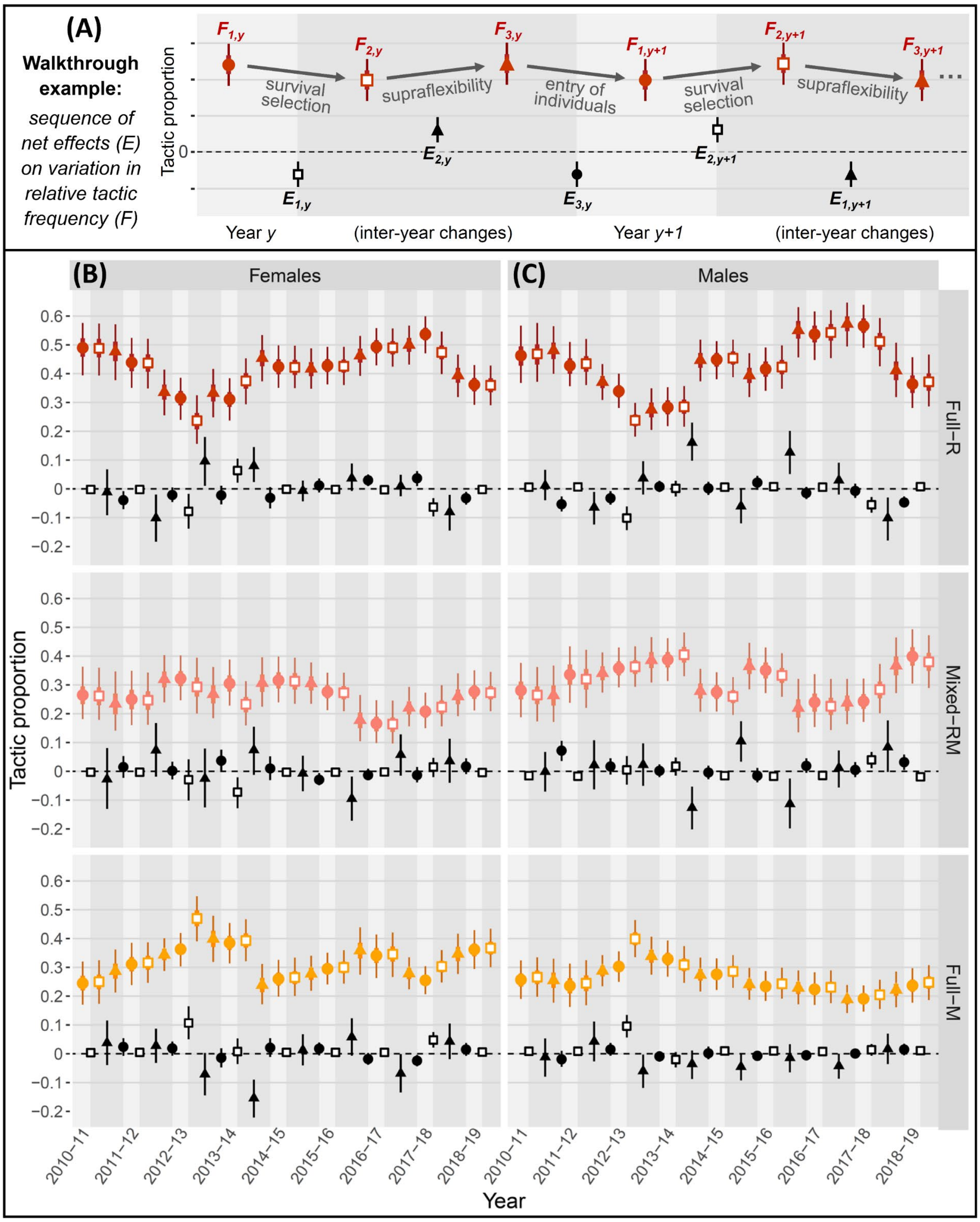


**Figure 7**. Overall synthesis of phenotypic dynamics of the three defined migratory tactics. The walkthrough example (A) illustrates that relative tactic frequency (*F*) is quantified at the start of each year *y* once new individuals have entered the adult population ($F_{1,y}$, filled coloured circles), then immediately after annual survival selection ($F_{2,y}$, open coloured squares), and immediately after subsequent tactic switching between *y* and *y*+1 ($F_{3,y}$, filled coloured triangles). The corresponding net

effects on variation in tactic frequency ($E$) from $y$ to $y$+1, due to survival selection ($E_{1,y}=F_{2,y}-F_{1,y}$, open black squares), supraflexibility ($E_{2,y}=F_{3,y}-F_{2,y}$, black triangles), and entry of new individuals ($E_{3,y}=F_{1,y+1}-F_{3,y}$, circles) are negative when frequency decreases and positive when it decreases. Estimates are shown for (B) females and (C) males, for full residence ('full-R', red, top row); mixed residence-migration ('mixed-RM', rose, middle row); full migration ('full-M', orange, bottom row). Coloured points are posterior means of $F$, with inner and outer line segments indicating 50% and 95% credible intervals. Black points are posterior means of $E$, with lines indicating 90% credible interval. A 90% credible interval located below (or above) zero indicates that the probability of $E$ being negative (or positive) is >95%.