

CONCEPT

The opportunity for selection: A slippery concept in ecology and evolution

Thomas E. Reed^{1,2}  | Marcel E. Visser³  | Robin S. Waples⁴ 

¹School of Biological, Earth & Environmental Sciences, University College Cork, Distillery Fields, Cork, Ireland

²Environmental Research Institute, University College Cork, Cork, Ireland

³Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

⁴Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA

Correspondence

Thomas E. Reed
Email: treed@ucc.ie

Funding information

H2020 European Research Council, Grant/Award Number: ERC-2014-StG-639192-ALH; SFI ERC Support Award

Handling Editor: Albert Phillimore

Abstract

1. Natural selection can only occur if individuals differ in fitness. For this reason, the variance in relative fitness has been equated with the 'opportunity for selection' (I), which has a long, albeit somewhat controversial, history. In this paper we discuss the use/misuse of I and related metrics in evolutionary ecology.
2. The opportunity is only realised if some fraction of I is caused by trait variation. Thus, $I > 0$ does not imply that selection is occurring, as sometimes erroneously assumed, because all fitness variation could be independent of phenotype.
3. The selection intensity on any given trait cannot exceed \sqrt{I} , but this upper limit will never be reached because (a) stochastic factors always affect fitness, and (b) there might be multiple traits under selection.
4. The expected magnitude of the stochastic component of I is negatively correlated with mean fitness. Uncertainty in realised I is also larger when mean fitness or population/sample size are low. Variation in I across time or space thus can be dominated (or solely driven) by variation in the strength of demographic stochasticity.
5. We illustrate these points using simulations and empirical data from a population study on great tits *Parus major*. Our analysis shows that the scope for fecundity selection in the great tits is substantially higher when using annual number of recruits as the fitness measure, rather than fledglings or eggs, even after adjusting for the dependence of I on mean fitness. This suggests nonrandom survival of juveniles across families between life stages. Indeed, previous work on this population has shown that offspring recruitment is often nonrandom with respect to clutch size and laying date of parents, for example.
6. We conclude that one cannot make direct inferences about selection based on fitness data alone. However, examining variation in ΔI_F (the opportunity for fecundity selection adjusted for mean fitness) across life stages, years or environments can offer clues as to when/where fecundity selection might be strongest, which can be useful for research planning and experimental design.

KEYWORDS

demography, eco-evolutionary, reproductive success, sexual selection, traits, wild populations

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Animal Ecology* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

1 | INTRODUCTION

In a 1958 paper about natural selection in humans (Crow, 1958), James Crow showed that the squared coefficient of variation in absolute fitness W , equivalent to the variance in relative fitness w ($w = W/\bar{W}$), places an upper limit on the rate of evolutionary adaptation (see also O'Donald, 1970):

$$CV_W^2 = \sigma_W^2 / \bar{W}^2 = \sigma_w^2 = I. \quad (1)$$

Crow called this 'the index of total selection', which he denoted by the symbol I , but it subsequently became known as the 'opportunity for selection' (Arnold & Wade, 1984).

I is a relevant parameter in the context of eco-evolutionary dynamics and population responses to environmental change, insofar as it constrains the degree to which evolution can shift trait distributions and increase mean fitness \bar{W} (and hence population growth). However, $I > 0$ is a necessary but not sufficient condition for selection on traits other than fitness itself. Selection on a given trait only occurs when some of the variation in fitness is caused by variation in the trait (Figure 1). Demographic stochasticity can account for a large and variable fraction of I (van Daalen & Caswell, 2019), so making inferences about the strength or drivers of selection based on the variance in fitness alone is problematic. This wider point has been made before in the context of sexual selection (Jennions et al., 2012), but the arguments and subtleties—which apply to natural selection in general—might not be familiar to ecologists. Our goal in this paper is therefore to guide empiricists on these issues. We first briefly review evolutionary theory on I . We then use simulations and empirical data on great tits to illustrate challenges relating to the interpretation of I and related metrics, and finish by offering some take-home messages.

From an evolutionary perspective, I is best measured via lifetime fitness (e.g. number of new-borns produced per new-born across its lifetime), as this will account for life-history trade-offs. However, I can be computed for any fitness component (e.g. survival, mating success, fecundity; Equation (1) applies to any fitness variable W) and is then interpreted as the scope for selection via that component

alone. We concentrate on survival and annual reproductive success in our examples, but for more involved treatments that partition variance in lifetime reproductive success in various ways, see (Arnold & Wade, 1984; Waples, 2022a; Waples & Reed, 2022), and references therein.

2 | THEORETICAL INTERPRETATION OF I

I is an important parameter in evolutionary biology because the selection intensity on any given character cannot exceed \sqrt{I} (Arnold & Wade, 1984), while the rate of evolution of fitness itself cannot exceed I (Crow, 1958; O'Donald, 1970; van Daalen & Caswell, 2019). Traits affecting fitness can then also evolve if they are heritable. The selection differential S on a focal trait Z is defined as the covariance between w and Z , or equivalently as the mean of the trait values weighted by w , minus the mean unweighted trait value:

$$S = \sigma_{Z,w} = \bar{Zw} - \bar{Z}. \quad (2)$$

The selection intensity i_z is then simply the selection differential expressed in trait standard deviation units: $i_z = S/\sigma_z$. This can also be written as $i_z = \rho_{Z,w}\sqrt{I}$, where $\rho_{Z,w}$ is the correlation between trait and fitness. Because the absolute value of this correlation is constrained to be ≤ 1 , this means that $|i_z| \leq \sqrt{I}$. The upper bound of \sqrt{I} will never be met, however, for at least three reasons. First, stochastic variation will explain some portion of fitness variation. Second, multiple independent traits might be under selection, so each will only explain a fraction of the remainder. Third, the relationship between any given trait and fitness might be nonlinear (Figure 1c), such that the linear correlation between them is less than 1 ($|\rho_{Z,w}| < 1$) even if fitness is an entirely deterministic function of this one trait (M. Morrissey, pers. comm.).

When n traits affect fitness, a multiple regression equation can be used to describe their independent linear effects; that is, $w = 1 + \beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_n Z_n + \varepsilon$, where the β coefficients correspond to selection gradients and ε is a noise term (Lande & Arnold, 1983). Nonlinear terms can be added but are ignored here for simplicity. i_z for each trait can then be expressed as the sum of

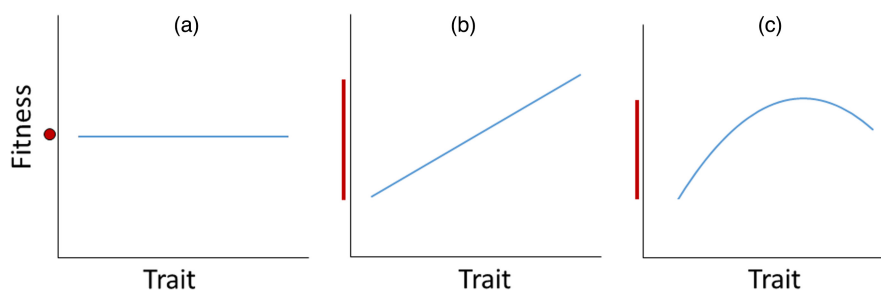


FIGURE 1 Phenotypic selection is associated with increased variation in fitness relative to a baseline scenario (a) where there is no relationship between trait and fitness (red dot indicates that expected fitness is constant). The second scenario (b) shows positive linear selection, with the red bar indicating the extent of trait-caused variation in expected fitness (which would be the same if the sign of the slope was reversed). On top of a linear component, there might be a nonlinear component to selection (c). The linear correlation between trait and expected fitness is then < 1 . Demographic stochasticity and/or independent selection on other traits adds random scatter to these relationships (not shown here for simplicity).

a direct component plus one or more indirect components owing to phenotypic correlations with other traits. For example, with two traits we have: $i_{z1} = \beta'_1 + \beta'_2 \rho_{1,2}$ and $i_{z2} = \beta'_2 + \beta'_1 \rho_{2,1}$ (where β'_1 and β'_2 are variance-standardised selection gradients; Lande & Arnold, 1983). Note that unlike $|i_z|$, $|\beta'_z|$ can exceed \sqrt{I} because the indirect component of selection can be opposite in sign to the direct component. The statistical contribution of each trait to the overall variance in relative fitness is then given by the product $i_z \beta'_z$, which equals i_z^2 when indirect selection is absent (Moorad & Wade, 2013).

Two lines of evidence suggest the lion's share of I might be attributable to stochastic variation in fitness. First, quantitative genetic studies have estimated that the heritability of fitness $h^2(w)$ (fraction of I explained by additive genetic effects) is low, typically on the order of a few per cent (Bonnet et al., 2022; Hendry et al., 2018). Second, even if many traits are under selection, they collectively might still explain relatively little of I . Assume, for argument's sake, that 10 independent traits are under linear selection and $i_z = 0.16$ for each (the median selection intensity reported by Kingsolver et al., 2001). The total contribution to I is then $10 \cdot (0.16^2) = 0.26$; overall I is then this plus some stochastic component. The stochastic component will

vary within and across studies, but one way of estimating the typical magnitude of overall I is to make use of the equation $I = \frac{I_A(w)}{h^2(w)}$, where $I_A(w)$ is the additive genetic variance in relative fitness, equivalent to the evolvability of fitness (Hansen & Houle, 2008). Using the median values for $I_A(w)$ and $h^2(w)$ of 0.10 and 0.03, respectively, reported by Bonnet et al. (2022), we have $I = \frac{I_A(w)}{h^2(w)} = \frac{0.10}{0.03} = 3.33$. Our 10 traits would thus together explain only 8% ($100 \cdot \frac{0.26}{3.33}$) of I , the rest being attributable to stochastic variation in fitness. This toy example of course makes a lot of assumptions, but it illustrates how stochastic effects can dominate I even with 10 traits under moderate selection.

3 | THE DEPENDENCE OF I ON MEAN FITNESS

A major limitation when comparing I across different contexts (populations, years, fitness variables) is that it is highly sensitive to mean fitness \bar{W} (Downhower et al., 1987), as \bar{W}^2 features in its denominator (Equation (1); Figure 2). No selection might occur on any trait in any environment, yet I can still vary simply because \bar{W} and/

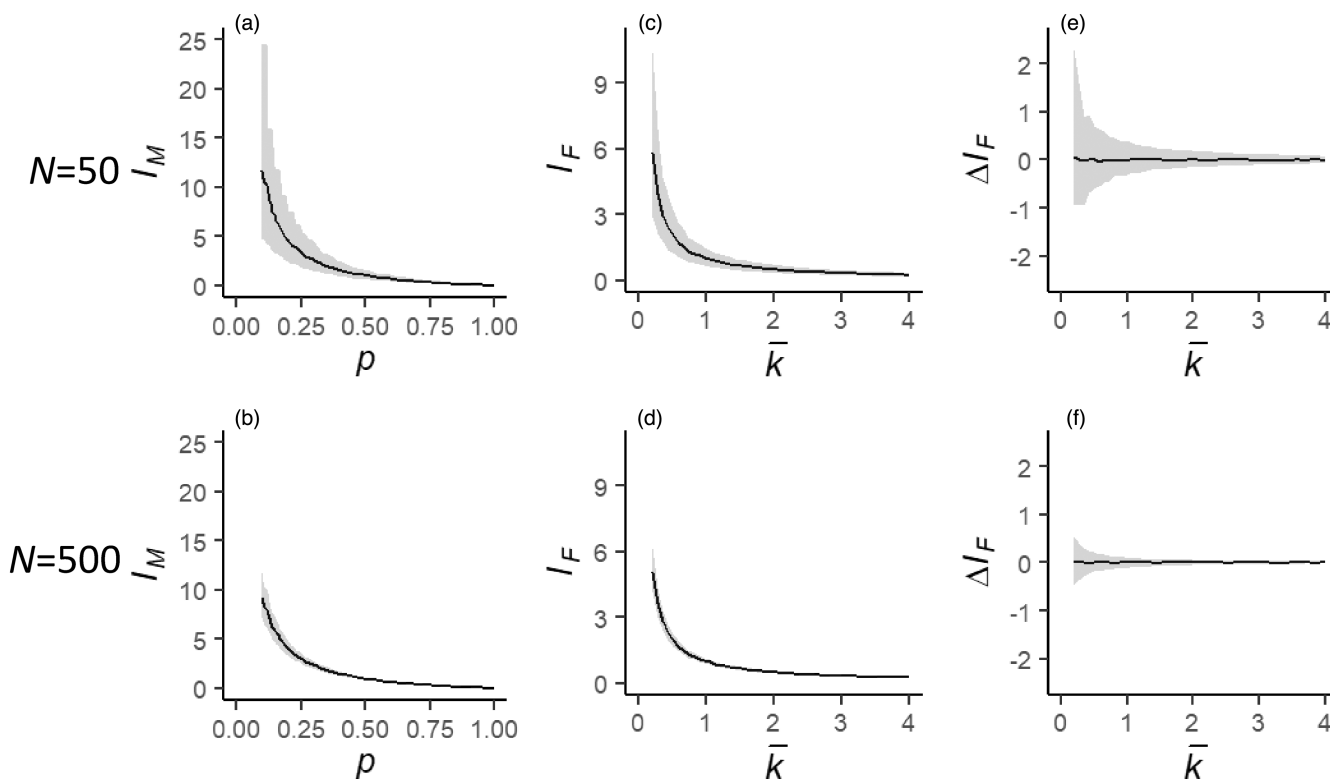


FIGURE 2 The opportunity for selection is inversely related to mean fitness, and uncertainty in I is higher when mean fitness is lower, or population size N is smaller. Each panel shows results of null simulations in which all fitness variation was assumed to be stochastic. Panels (a) and (b): The opportunity for viability selection I_M plotted as a function of the survival rate p . For each of 50 different values of p , $N = 50$ (a) or $N = 500$ (b) individuals were simulated, where each had the same expected survival. Realised survival was determined by drawing a random number from a standard uniform distribution and killing the individual if this number was $< p$. This was repeated 10,000 times to generate approximate 95% confidence intervals (grey bands) for I_M . Panels (c–f): Results of simulations in which the number of offspring k for each of $N = 50$ (c, e) or $N = 500$ (d, f) parents was simulated by drawing from a Poisson distribution of mean and variance equal to \bar{k} . 10,000 replicate simulations were performed for each of 50 different \bar{k} values, to generate approximate 95% confidence intervals for the opportunity for fecundity selection I_F (C,D) and the adjusted metric, ΔI_F , of Waples (2020). Black curves in all panels are mean values across replicate simulations.

or σ_W^2 (demographic stochasticity) varies. If selection does occur, it can be stronger in harsher, or human-disturbed, environments where \bar{W} is lower (Fugère & Hendry, 2018; Hunter et al., 2018; Reiss, 2013), because \bar{W} features in the denominator of the selection differential (which can be reformulated in absolute fitness terms as $S = \text{cov}(W, Z) / \bar{W}$). Stronger selection is not guaranteed at lower \bar{W} , however, because $\text{cov}(W, Z)$ might also be lower when \bar{W} is lower (e.g. if fitness reductions are proportional across individuals; Fugère & Hendry, 2018). In general, ecological changes can affect S via changes in the mean or variance of the trait distribution, or via changes in the intercept or slope of the trait-absolute fitness relationship (Hunter et al., 2018).

Even if selection on a given trait varies in intensity as \bar{W} changes, I and i_z might still be only weakly coupled. Multiple traits will typically be under selection, and if selection intensities on each are uncorrelated (or weakly correlated) across environments this effectively

adds noise to the expected positive relationship (Figure 3) between I and i_z^2 for any given trait. The stochastic component of I will also be highly variable in magnitude. Its expected value will be some function of \bar{W} , owing to mean-variance scaling for non-Gaussian fitness variables such as survival and reproductive success (Figure 2). Its realised magnitude, for a given \bar{W} , will also vary randomly around this expectation (Section 4; Figure 2). Systematic and/or random variation in the magnitude of demographic stochasticity thus render relationships between I and explanatory variables of interest vulnerable to misinterpretation (Jennions et al., 2012).

With binary fitness variables like survival, the variance equals $p(1-p)$, where p is the survival rate. The opportunity for viability selection (I_M , where the m stands for mortality) is then given as $I_M = \frac{p(1-p)}{p^2} = \frac{1-p}{p}$ (Crow, 1958). For a given p , I_M will always be the same, regardless of whether there is a nonrandom (e.g. trait-determined) component to mortality or not, because latent variation

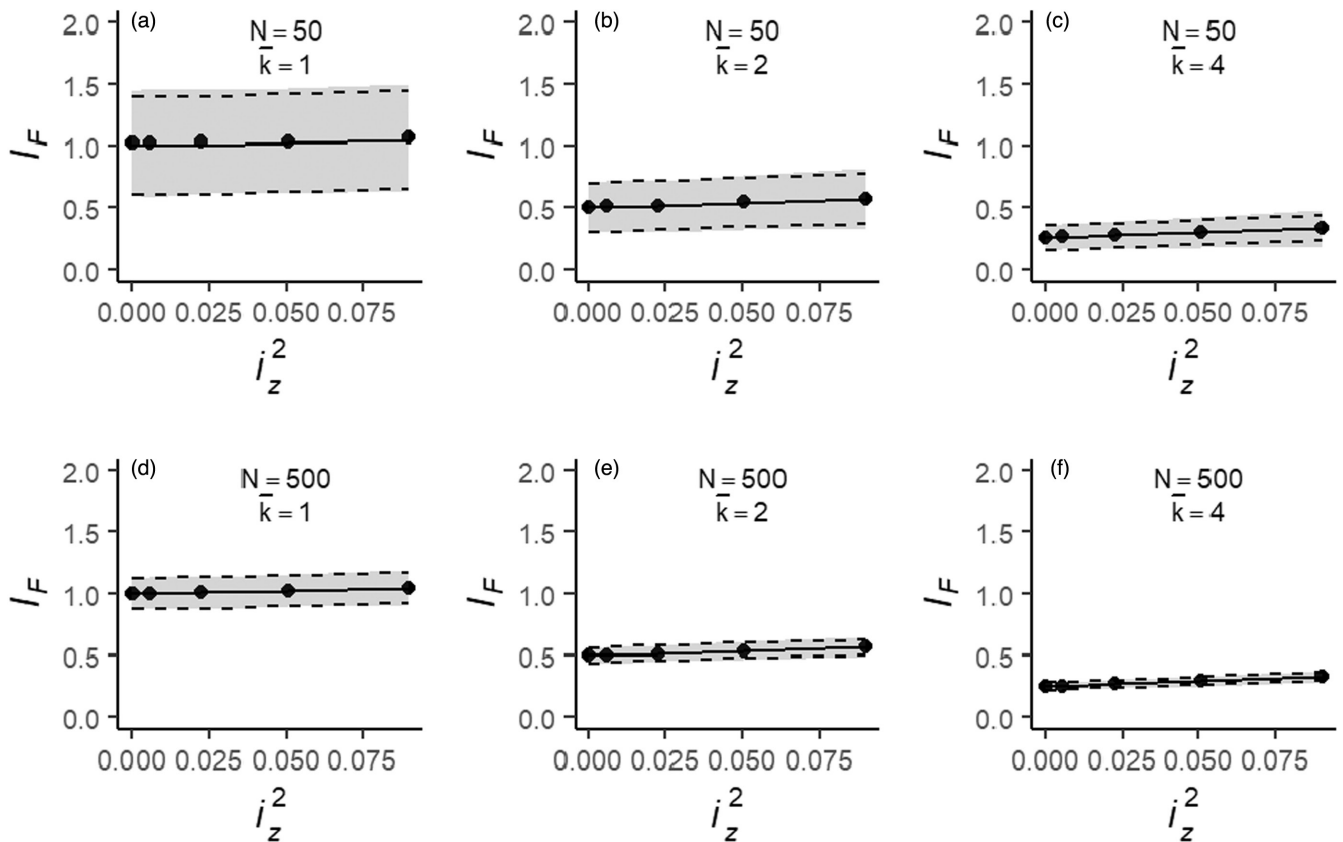


FIGURE 3 Simulations illustrating the relationship between I_F and i_z^2 (squared selection intensity). The logarithm of the expected number of offspring W for each of N parents was defined as a linear function of an arbitrary parental trait Z : $\ln(W) = a + bZ$. This was converted to realised integer number of offspring k by drawing from a Poisson distribution: $k \sim \text{pois}(W)$. I_F was then computed as the variance in k divided by \bar{k}^2 . Six scenarios were explored by setting $N = 50$ (panel a–c) or $N = 500$ (d–f) and $\bar{k} = 1$ (a, d), $\bar{k} = 2$ (b, e) or $\bar{k} = 4$ (c, f). In each scenario, five strengths of selection were simulated by varying b from 0 to 0.3 in steps of 0.075. Note that b (the log-scale slope of the fitness function) corresponds to the selection gradient (Morrisey & Goudie *Evolution*) and in this case also to i_z , as Z was drawn from a normal distribution of zero mean and unit variance. The intercept of the fitness function in each scenario was fixed at the natural logarithm of the desired \bar{k} . Mean realised I (filled dots) and its confidence intervals (grey bands) across 10,000 replicate simulations per selection strength were then calculated. Solid line = theoretically expected I , which is the sum of the trait-determined component (i_z^2) and the expected stochastic component ($1/\bar{k}$ in this case, because a Poisson null model was assumed; note that assuming over- or underdispersion in random reproductive success would simply shift the intercepts of the relationships in each plot). Dashed lines = approximate theoretical confidence intervals on I (lower = $i^2 - 2\sqrt{2/(N\bar{k}^2)}$; upper = $i^2 + 2\sqrt{2/(N\bar{k}^2)}$; see Appendix S1).

in a Bernoulli variable is unobservable. If all one has is data on survival, there is no way to correct I_M for its dependence on mean survival (Figure 2) to try to get at the nonrandom component.

The opportunity for fecundity selection I_F can in theory be corrected for its dependence on (population or sample) mean fecundity. Let k be the number of offspring produced per parent in a single reproductive bout, or across a full breeding season, that survive to the age at which juveniles are counted. An adjusted I_F can then be computed as $\Delta I_F = I_F - 1/\bar{k}$ (Waples, 2020). Note that here we are defining ‘fecundity’ broadly as the product of zygote number and offspring survival to the stage at enumeration. We acknowledge that assigning offspring survival as a component of parental fitness is problematic from a quantitative genetic theory perspective (Thomson & Hadfield, 2017) but argue that computing ΔI_F for such ‘mixed fitness’ measures still is useful as an overall index of potential for selection to be acting. Under a Wright–Fisher model—a mainstay of classical population genetics— k is approximately Poisson distributed when reproductive success is purely random (i.e. $\sigma_k^2 \approx \bar{k}$). The expected null value of I_F is then $1/\bar{k}$, so subtracting this quantity from raw I_F not only corrects for the dependence on \bar{k} , but also for the magnitude of demographic stochasticity expected under Wright–Fisher reproduction, under which $E(\Delta I_F) = 0$ (Waples, 2020). Thus positive ΔI_F values suggest, but do not prove, that something interesting might be going on. Assuming a Poisson distribution in the null case will not always be warranted: k might be underdispersed ($\sigma_k^2 < \bar{k}$) or overdispersed ($\sigma_k^2 > \bar{k}$) in some species/populations even if reproductive success is purely random (Kendall & Wittmann, 2010; Waples & Reed, 2022). Thus, negative, or positive, ΔI_F values might be expected absent any selection, and deviations from a given expectation can also occur by chance (Figure 2; Section 4).

Given these limitations, interpreting absolute values of ΔI_F is problematic, but comparing relative values across contexts where \bar{k} varies can be informative. Regardless of the magnitude of overdispersion expected under a given null model, raw I_F is expected to be higher when \bar{k} is lower, even if selection is always absent. The same is not true of ΔI_F , however. If the overdispersion parameter φ ($\varphi = \sigma_k^2 / \bar{k}$) varies across environments, as expected if selection intensities vary, $E(\Delta I_F)$ will in turn also vary (but note it might be nonzero even in environments where selection is absent). Changes in ΔI_F therefore more reliably indicate changes in the true opportunity for selection than do changes in raw I_F (Table 1). Nevertheless, there remains no free lunch. The only way to show that selection intensities on a given trait of interest do in fact vary is to test for variation in the slope of the relationship between trait and relative fitness (Morrissey & Hadfield, 2012; Wade & Kalisz, 1990), which obviously requires phenotypic information as well as fitness information.

Above, we defined k as the number of offspring per parent surviving up to a particular point. The later in life offspring are counted, the lower \bar{k} will be, and thus the higher raw I_F will be. Consider a highly fecund species such as cod that produces hundreds of thousands of eggs. For a given level of real fitness differences among parents, I_F computed via number of eggs will be tiny, whereas I_F computed via number of juveniles surviving a full year will be much larger. On the one hand, this makes biological sense, in that the more offspring mortality that has accrued, the higher the maximum possible selection intensities on parental traits influencing offspring survival. On the other hand, offspring survival might be purely random with respect to parental phenotype, in which case the true scope for selection does not actually increase just because offspring are enumerated at older ages.

TABLE 1 Numerical example based on hypothetical reproductive success data. The opportunity for fecundity selection, I_F , is higher when juveniles are counted at a later life stage (stage B) at which both the mean (\bar{k}) and variance (σ_k^2) in offspring number is lower, relative to stage A. This increase in I_F occurs regardless of whether the intervening offspring mortality is completely random (scenario 1) or not (scenario 2). In contrast, ΔI_F only increases in scenario 2, reflecting a real increase in the (adjusted) opportunity for selection. In both scenarios, we assume arbitrarily that there is overdispersion at stage A, but we make no assumptions regarding whether this stems from heritable or nonheritable sources of variation. Random offspring mortality between stages A and B shrinks any initial overdispersion ($\varphi > 1$) towards the Poisson expectation of $\varphi = 1$. This shrinkage is less pronounced in scenario 2, as offspring mortality now has a nonrandom component. Also shown is the scaled overdispersion parameter of Crow & Morton, 1955, computed as $\varphi' = 1 + \frac{\bar{k}_B}{\bar{k}_A}(\varphi_A - 1)$, which rescales φ_A to its expected value given observed \bar{k} at life stage B assuming intervening mortality is random. The most informative comparison (highlighted in bold) is between φ' at stage A and observed φ at stage B. In scenario 1, $\varphi_B = \varphi'_A$, so there is no evidence for selection between stages. In scenario 2, however, $\varphi_B > \varphi'_A$, indicating nonrandom survival (which might be due to selection, but phenotypic information is then required to test this). Bootstrapping could be used to test whether a given difference between φ_B and φ'_A (or between $\Delta I_{F(B)}$ and $\Delta I_{F(A)}$) is statistically significant (see main text)

Parameter	Symbol	Scenario 1 (random mortality)		Scenario 2 (nonrandom mortality)	
		Juvenile stage A	Juvenile stage B	Juvenile stage A	Juvenile stage B
Mean offspring number	\bar{k}	10	2	10	2
Variance offspring number	σ_k^2	15	2.2	15	2.8
Overdispersion parameter	φ	1.5	1.1	1.5	1.4
Scaled overdispersion par.	φ'	1.1	—	1.1	—
Opportunity for selection	I_F	0.15	0.55	0.15	0.70
Adjusted opportunity for selection	ΔI_F	0.05	0.05	0.05	0.10

Again, ΔI_F offers some advantages over raw I_F : comparing ΔI_F across consecutive offspring life stages can provide insights as to whether any intervening mortality is random or not (Table 1). An increase in ΔI_F above that expected by chance (see Section 5) suggests that selection *might* be occurring in the intervening period, but equally it might have nothing to do with the phenotype of parents (or offspring). For example, offspring from the same parents will co-occur in space for some period in ontogeny (e.g. siblings sharing a nest or parental territory), so entire families can live or die as units owing to random catastrophes or variation in habitat quality.

As well as being sensitive to population (i.e. true) \bar{k} , raw I_F is also sensitive to sample \bar{k} . Owing to constraints on study design, it might be possible to sample only a random subset of all offspring alive at a given age/stage. If sampling effort varies, this will produce variation in \bar{k} estimates that has nothing to do with biology. Randomly sampling offspring at an early life stage is statistically equivalent to enumerating all offspring at some later life stage after which random mortality has occurred (as in Table 1, Scenario 1). $E(\Delta I_F)$ at this early life stage will be the same regardless of sampling effort, but the same is not true for $E(I_F)$. Sampling a smaller fraction of offspring will bias I_F upwards (because sample \bar{k} is lower relative to true \bar{k}), giving the false impression of more scope for selection, whereas ΔI_F is unbiased in this regard. Note that if trait information is available, phenotypic selection estimates are not biased by lower sampling effort, but their uncertainty would be higher.

4 | UNCERTAINTY IN I

For a given mean fitness, the realised value of I will vary randomly around some expectation owing to sampling effects. Uncertainty in the stochastic component of I is higher when \bar{W} is lower (Figure 2). For a given \bar{W} , uncertainty is also higher when population or sample size is lower. When selection occurs on one or more traits, sampling variation around the true selection intensities (i.e. population values of i_z) will in turn be higher when \bar{W} and/or sample size is lower. In other words, phenotypic selection is itself a stochastic process, and the realised magnitude of the trait-determined component of I in any given environment will depend on the realised selection pressures.

To illustrate these points, we simulate a simple scenario where a single trait Z is under linear selection via variation in offspring number k . One way to model this is to work with expected fitness W on the natural logarithm scale (Morrissey & Goudie, 2016), such that negative values are impossible: $\ln(W) = a + bZ$, where a is the intercept and b the slope of the individual fitness function. Expected absolute number of offspring is then given by $W = \exp(a + bZ)$, and this can be converted into a random variable by assuming some stochastic process. Here, for simplicity, we assume a Poisson process ($k \sim \text{pois}(W)$), but any distribution could be used. This is conceptually equivalent to simulating a generalised Wright–Fisher model with weights given by W (Waples, 2022b). A range of selection intensities was produced by varying b across simulations, holding a constant at

the log of the desired \bar{W} such that $E(\bar{k})$ was independent of selection strength (as might occur for example with soft selection; Bell et al., 2021).

As expected, the results showed that realised I_F increased with i_z^2 (Figure 3). This relationship was always linear because $E(\bar{k})$ was constant within each scenario. Curved relationships would have instead resulted if $E(\bar{k})$ varied within scenarios (e.g. by assuming some relationship between the a and b parameters), because the trait-determined and stochastic components of I_F would then no longer be independent. Uncertainty in I_F across replicate simulations was larger when N was smaller or \bar{k} was lower. The realised correlation between I_F and i_z^2 for a given replicate was thus lower and more variable when N or \bar{k} was lower. When selection was absent ($i_z = 0$), the sampling variance in I_F (and ΔI_F) was well approximated by $2/N\bar{k}^2$ (Appendix S1). With real-world data on some fitness variable, confidence intervals around estimates of I_F or ΔI_F could be computed via bootstrapping (see Section 5).

5 | ILLUSTRATION OF EMPIRICAL PATTERNS WITH GREAT TIT DATA

To illustrate key points made in previous sections, we exploit 46 years (1973 to 2018) of individual-level reproductive success data from a Dutch study population (National Park de Hoge Veluwe, the Netherlands) of great tits *Parus major* (small songbirds common across Eurasia) see Visser et al. (2021) and references therein for details on the study population and data collection methods. Annual reproductive output k per known (ringed) female was measured in three different ways: number of eggs (including second clutches but excluding replacement clutches), number of fledglings (juveniles surviving the nestling phase; all clutches included) and number of recruits (first-time breeders recorded in the study area the following years; all clutches included). Birds that underwent experimental manipulations were excluded. Average values (across all females and years combined) for the same parameters listed in Table 1 were then calculated. Raw I_F was an order of magnitude greater for recruits compared to fledglings, which was in turn about twice that for eggs (Table 2). This simply reflects the fact that the biggest drop in \bar{k} was between fledglings and recruits, given the substantial intervening mortality. Overdispersion ($\phi > 1$) was apparent for all three fitness variables, particularly fledglings (Table 2). Note, however, that clutch/litter size is often underdispersed in other species (Kendall & Wittmann, 2010) and egg number might be underdispersed in other great tit populations, for example, those where second clutches are uncommon. The scaled overdispersion parameter (ϕ') of Crow and Morton (1955) was close to 1 for eggs and fledglings, indicating that random mortality up to the recruit stage would shrink initial overdispersion (or grow initial underdispersion) towards the Poisson expectation of 1. However, unscaled ϕ was 1.54 at the recruit stage, implying that intervening mortality was not completely random. Consistent with this, ΔI_F was much higher for recruits, compared to fledglings or eggs.

TABLE 2 Parameter estimates for the great tit data. See [Table 1](#) legend and main text for parameter definitions. Mean and variance in offspring number estimated across all females and years combined, and other parameters calculated from these. Result very similar when parameters instead estimated within years and then annual values averaged across years. φ' for eggs and fledglings corresponds to their expected φ values at the recruit stage (given observed mean number of recruits) if intervening mortality was random

Parameter	Symbol	Eggs	Fledged	Recruits
Mean offspring number	\bar{k}	9.76	6.24	0.44
Variance offspring number	σ_k^2	13.33	13.13	0.67
Overdispersion parameter	φ	1.37	2.10	1.54
Scaled overdispersion par.	φ'	1.02	1.08	—
Opportunity for selection	I_F	0.14	0.34	3.53
Adjusted opportunity for selection	ΔI_F	0.04	0.18	1.24

To assess whether the changes in ΔI_F between stages were more than expected by chance (bootstrapped 95% confidence intervals (BCIs) nonoverlapping zero), a nonparametric bootstrapping procedure was performed in which $n = 100$ 'females-years' (an instance of a particular female breeding in a particular year; some females breed across multiple years) were randomly sampled, with replacement, from all 2486 female-year records. The difference in ΔI_F between each pair of stages (eggs-fledglings; fledglings-recruits; eggs-recruits) was then computed for this random sample, and this was repeated 10,000 times to generate bootstrapped distributions. The observed increase in ΔI_F between eggs and fledglings of 0.14 was more than expected by chance (lower BCI = 0.06; upper BCI = 0.23). The same was true for the difference in ΔI_F between fledglings and recruits (observed = 1.07; lower BCI = 0.14; upper BCI = 2.20), and for the difference in ΔI_F between eggs and recruits (observed = 1.20; lower BCI = 0.28; upper BCI = 2.35).

Supporting evidence for a nonrandom component to reproductive success in this study population comes from the fact that annual number of recruits is repeatable across females (Reed et al., 2016). Laying date and clutch size are two heritable traits that explain some of this individual heterogeneity among females, with selection on each trait varying in intensity and to some extent sign across years (Reed et al., 2016; Sæther et al., 2016; Visser et al., 2021). Additional traits might be under selection, and the big jump in ΔI_F between fledglings and recruits ([Table 2](#)) suggests that the combined fitness effects of all traits are strongest during the postfledgling period. Environmental factors shared by families can also increase ΔI_F above the null expectation, but such 'common environment effects' are expected to be stronger during the nestling phase (when siblings are still together) than in the postfledgling phase. Across years, the correlation between I_F and i_z^2 for laying date (both estimated via number of recruits) was weak (Spearman's $r = 0.26$) and not statistically significant ($p = 0.08$). Similarly, the correlation between I_F and i_z^2 for clutch size (both again estimated via number of recruits) was weak and not significant (Spearman's $r = -0.08$; $p = 0.61$). This suggests that the realised magnitude of demographic stochasticity varies substantially across years, and/or that selection on other unmeasured traits varies in intensity. Either way, the component of I_F not explained by laying date or clutch size appears to be highly variable across years.

6 | ETHICS/LICENCING

The research was carried out under licence AVD801002017831 of the Centrale Commissie Dierexperimenten (CCD) in the Netherlands. Fieldwork at the National Park de Hoge Veluwe was carried out with permission of the Park.

7 | DISCUSSION

I remains theoretically relevant in population biology, in that it places a hard ceiling on the rate of evolutionary adaptation. The more salient parameter, although, is $I_A(w)$, as this captures the *actual* rate of evolutionary adaptation, that is, the genetic increase in mean fitness due to selection (Bonnet et al., 2022; Hendry et al., 2018). Given that most variation in fitness typically is nonheritable, and that the heritability of fitness ($h^2(w) = I_A(w) / I$) might vary unpredictably across environments, I is of limited practical utility for understanding/predicting evolutionary dynamics (Grafen, 1988). At best, it provides an answer to the question: 'What is the maximum amount by which evolution could increase population growth in this environment?', which might help guide management/conservation scenario-planning, for example. I also provides an answer to the question: 'How strong could selection be on any trait?'. The related OSM metric of Pelletier and Coulson (2012) in turn provides an answer to the more specific question: 'How strong could selection be on *my particular study trait*'. Again, however, knowing how strong things *could be* is of less practical utility than knowing how strong they in fact are.

More fundamentally, $I > 0$ simply cannot be taken as evidence for selection, as all fitness variation could be random with respect to (multivariate) phenotype. Variation in I across ecological contexts also cannot be taken as evidence that selection is varying, as the only thing that might be varying is the magnitude of demographic stochasticity. If mean fitness varies, the stochastic component of I will then vary systematically, so realised I is then an even-less reliable guide to selection. Similarly, if demographic stochasticity varies systematically with some putative environmental driver of selection, the scope for being misled is high (Jennions et al., 2012). ΔI_F corrects for dependence on mean offspring number but is also open to misinterpretation (Waples, 2020; Waples & Reed, 2022). Comparing

relative values of ΔI_F across situations where true or sample \bar{k} varies can nonetheless give a more reliable indication of the true scope for selection than comparing raw I_F . Our empirical example with the great tits—a system in which parental reproductive success is known to be nonrandom with respect to laying date and clutch size—supported this latter contention.

Our simulations showed that the realised correlation between I and the squared selection intensity on a given trait can be weak and highly variable, especially when mean fitness or population/sample size are low. Indeed, with the great tits, I_F was statistically uncorrelated with i_z^2 for both laying date and clutch size, despite 46 years of data. Few other studies seem to have explored this, but a study of water striders *Aquarius remigis* found a weak positive correlation ($r_s = 0.34$, $n = 24$) between I and standardised linear selection gradients on body size (Ferguson & Fairbairn, 2001). The data were quite heterogeneous, however, involving three different fitness variables, complicating the comparison with the great tit results. Pelletier and Coulson (2012) documented positive relationships between viability selection differentials on juvenile body size and I_M in both red deer *Cervus elaphus* and Soay sheep *Ovis aries*, but the R^2 values (0.30 and 0.15, respectively, including outliers) were still relatively low (see also Martin et al., 2015).

In conclusion, we hope our arguments and examples have clarified the dangers of uncritically using I and related metrics to make inferences about selection and its drivers. There is simply no free lunch: measuring selection requires data on phenotypes! Nevertheless, comparing I across life stages or environments can provide hints as to where selection might be acting most strongly, with ΔI_F being preferable to raw I_F when \bar{k} varies. This might in turn prompt further study or inform study design. If one does have phenotypic information from multiple years/locations, on top of paired fitness data, correlating squared selection intensities against I can still be revealing. For example, if mean fitness is high and relatively constant, and population size is large, then variability in the stochastic component of I should be low. A low correlation between I and i_z^2 for a focal trait would then imply variable selection intensities on one or more unmeasured traits. Wider reporting of the mean and variance of absolute fitness (by year/location), as well as the mean and variance of any measured traits and their relationship with absolute fitness (Hunter et al., 2018), will allow for improved general understanding of the ecological drivers of selection.

AUTHOR CONTRIBUTIONS

All authors contributed intellectually to the paper. The paper was written by Thomas E. Reed, with inputs from Robin S. Waples and Marcel E. Visser.

ACKNOWLEDGEMENTS

Thanks to Jacob Moorad for discussions on the OFS, and to Joel Pick, Michael Morrissey and an anonymous reviewer for thoughtful and detailed reviews. Thanks to Maria Teider and Roy Supratik for help on mathematical aspects. We also thank the many workers who

contributed to the great tit data collection and collation over the years and the National Park de Hoge Veluwe for permission to work on their premises. TER was funded by an ERC Starting Grant (ERC-2014-StG-256 639192-ALH) and an SFI ERC Support Award. Open access funding provided by IReL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.tmpg4f52h> (Reed et al., 2022).

ORCID

Thomas E. Reed  <https://orcid.org/0000-0002-2993-0477>

Marcel E. Visser  <https://orcid.org/0000-0002-1456-1939>

Robin S. Waples  <https://orcid.org/0000-0003-3362-7590>

REFERENCES

- Arnold, S. J., & Wade, M. J. (1984). On the measurement of natural and sexual selection: Theory. *Evolution*, 38(4), 709–719.
- Bell, D. A., Kovach, R. P., Robinson, Z. L., Whiteley, A. R., & Reed, T. E. (2021). The ecological causes and consequences of hard and soft selection. *Ecology Letters*, 24(7), 1505–1521.
- Bonnet, T., Morrissey, M. B., de Villemereuil, P., Albers, S. C., Arcese, P., Bailey, L. D., Boutin, S., Brekke, P., Brent, L. J., & Camenisch, G. (2022). Genetic variance in fitness indicates rapid contemporary adaptive evolution in wild animals. *Science*, 376(6596), 1012–1016.
- Crow, J. F. (1958). Some possibilities for measuring selection intensities in man. *Human Biology*, 30(1), 1–13.
- Crow, J. F., & Morton, N. E. (1955). Measurement of gene frequency drift in small populations. *Evolution*, 9, 202–214.
- Downhower, J. F., Blumer, L. S., & Brown, L. (1987). Opportunity for selection: An appropriate measure for evaluating variation in the potential for selection? *Evolution*, 41(6), 1395–1400.
- Ferguson, I. M., & Fairbairn, D. J. (2001). Is selection ready when opportunity knocks? *Evolutionary Ecology Research*, 3(2), 199–207.
- Fugère, V., & Hendry, A. P. (2018). Human influences on the strength of phenotypic selection. *Proceedings of the National Academy of Sciences of the United States of America*, 115(40), 10070–10075.
- Grafen, A. (1988). On the uses of data on lifetime reproductive success. In T. H. Clutton-Brock (Ed.), *Reproductive success* (pp. 454–471). Chicago University Press.
- Hansen, T. F., & Houle, D. (2008). Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology*, 21(5), 1201–1219.
- Hendry, A. P., Schoen, D. J., Wolak, M. E., & Reid, J. M. (2018). The contemporary evolution of fitness. *Annual Review of Ecology, Evolution, and Systematics*, 49, 457–476.
- Hunter, D. C., Pemberton, J. M., Pilkington, J. G., & Morrissey, M. B. (2018). Quantification and decomposition of environment-selection relationships. *Evolution*, 72(4), 851–866.
- Jennions, M. D., Kokko, H., & Klug, H. (2012). The opportunity to be misled in studies of sexual selection. *Journal of Evolutionary Biology*, 25(3), 591–598.
- Kendall, B. E., & Wittmann, M. E. (2010). A stochastic model for annual reproductive success. *The American Naturalist*, 175(4), 461–468.
- Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P., & Beerli, P. (2001). The strength

- of phenotypic selection in natural populations. *The American Naturalist*, 157(3), 245–261.
- Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, 37, 1210–1226.
- Martin, A. M., Festa-Bianchet, M., Coltman, D. W., & Pelletier, F. (2015). Comparing measures of breeding inequality and opportunity for selection with sexual selection on a quantitative character in bighorn rams. *Journal of Evolutionary Biology*, 28(1), 223–230.
- Moorad, J. A., & Wade, M. J. (2013). Selection gradients, the opportunity for selection, and the coefficient of determination. *The American Naturalist*, 181(3), 291–300.
- Morrissey, M. B., & Goudie, I. B. (2016). Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions. *Evolution*, 76(7), 1378–1390.
- Morrissey, M. B., & Hadfield, J. D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution: International Journal of Organic Evolution*, 66(2), 435–442.
- O'Donald, P. (1970). Change of fitness by selection for a quantitative character. *Theoretical Population Biology*, 1(2), 219–232.
- Pelletier, F., & Coulson, T. (2012). A new metric to calculate the opportunity for selection on quantitative characters. *Evolutionary Ecology Research*, 14(6), 729–742.
- Reed, T. E., Gienapp, P., & Visser, M. E. (2016). Testing for biases in selection on avian reproductive traits and partitioning direct and indirect selection using quantitative genetic models. *Evolution*, 70(10), 2211–2225.
- Reed, T. E., Visser, M. E., & Waples, R. S. (2022). Data from: The opportunity for selection: A slippery concept in ecology and evolution. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.tmpg4f52h>
- Reiss, J. O. (2013). Does selection intensity increase when populations decrease? Absolute fitness, relative fitness, and the opportunity for selection. *Evolutionary Ecology*, 27(3), 477–488.
- Sæther, B.-E., Visser, M. E., Grøtan, V., & Engen, S. (2016). Evidence for r- and K-selection in a wild bird population: A reciprocal link between ecology and evolution. *Proceedings of the Royal Society B: Biological Sciences*, 283(1829), 20152411.
- Thomson, C. E., & Hadfield, J. D. (2017). Measuring selection when parents and offspring interact. *Methods in Ecology and Evolution*, 8(6), 678–687.
- van Daalen, S., & Caswell, H. (2019). Demographic sources of variation in fitness. In R. Sear, R. Lee, & O. F. H. Burger (Eds.), *Human evolutionary demography*. Open Book Publishers.
- Visser, M. E., Lindner, M., Gienapp, P., Long, M. C., & Jenouvrier, S. (2021). Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (*Parus major*). *Proceedings of the Royal Society B: Biological Sciences*, 288(1963), 20211337.
- Wade, M. J., & Kalisz, S. (1990). The causes of natural selection. *Evolution*, 44(8), 1947–1955.
- Waples, R. S. (2020). An estimator of the opportunity for selection that is independent of mean fitness. *Evolution*, 74(9), 1942–1953.
- Waples, R. S. (2022a). Partitioning variance in reproductive success, within years and across lifetimes. *BioRxiv*. <https://www.biorxiv.org/content/10.1101/2022.02.08.479606v1>
- Waples, R. S. (2022b). THEWEIGHT: A simple and flexible algorithm for simulating non-ideal, age-structured populations. *Methods in Ecology and Evolution*, 13, 2030–2041. <https://doi.org/10.1111/2041-210X.13926>
- Waples, R. S., & Reed, T. E. (2022). Null models for the opportunity for selection. *BioRxiv*. <https://biorxiv.org/cgi/content/short/2022.06.30.498333v1>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Reed, T. E., Visser, M. E., & Waples, R. S. (2023). The opportunity for selection: A slippery concept in ecology and evolution. *Journal of Animal Ecology*, 92, 7–15. <https://doi.org/10.1111/1365-2656.13841>