Evolution, 2024, **XX(XX)**, 1–15
https://doi.org/10.1093/evolut/qpae043
Advance access publication 13 March 2024
Perspective

Fitness as the organismal performance measure guiding adaptive evolution

Lutz Fromhage[1], Michael D. Jennions[2,3], Lauri Myllymaa[1], Jonathan M. Henshaw[4]

[1]Department of Biological and Environmental Science, University of Jyvaskyla, Jyvaskyla, Finland
[2]Evolution & Ecology, Research School of Biology, Australian National University, Canberra, ACT, Australia
[3]Stellenbosch Institute for Advanced Study (STIAS), Wallenberg Research Centre at Stellenbosch University, Stellenbosch 7600, South Africa
[4]Institute of Biology I (Zoology), University of Freiburg, Freiburg, Germany

Corresponding author: Department of Biological and Environmental Science, University of Jyvaskyla, P.O. Box 35, 40014 Jyvaskyla, Finland.
Email: lutz.fromhage@jyu.fi

Abstract

A long-standing problem in evolutionary theory is to clarify in what sense (if any) natural selection cumulatively improves the design of organisms. Various concepts, such as *fitness* and *inclusive fitness*, have been proposed to resolve this problem. In addition, there have been attempts to replace the original problem with more tractable questions, such as whether a given gene or trait is favored by selection. Here, we ask what theoretical properties the concept *fitness* should possess to encapsulate the improvement criterion required to talk meaningfully about adaptive evolution. We argue that natural selection tends to shape phenotypes based on the causal properties of individuals and that this tendency is, therefore, best captured by a fitness concept that focuses on these properties. We highlight a fitness concept that meets this role under broad conditions but requires adjustments in our conceptual understanding of adaptive evolution. These adjustments combine elements of Dawkinsian gene selectionism and Egbert Leigh's "parliament of genes."

Keywords: natural selection, adaptation, social evolution, inclusive fitness, kin selection, causality

> *Now, as each of the parts of the body, like every other instrument, is for the sake of some purpose, viz. some action, it is evident that the body as a whole must exist for the sake of some complex action* (Aristotle, De Partibus Animalium, 1.5, 645b15–18).
> *Very often, a helpful way of thinking about the evolution of some behavioural trait is to imagine a gene of very low penetrance, and to ask, on those occasions when the gene is expressed (that is, when the trait appears) is the result an increase or a decrease in the number of copies of the gene?* (Maynard Smith, 1998, p. 137).

1. Introduction

In *The Origin of Species*, Darwin described the long-term effect of natural selection as "the accumulation of innumerable slight variations, each good for the original possessor" (Darwin, 1859, p. 459), thereby leading "to the improvement of each creature in relation to its organic and inorganic conditions of life" (Darwin, 1860, p. 127). The idea of an inherent tendency in natural selection toward cumulative improvement remains relevant today because natural selection is still our only scientific explanation for complex adaptive design in nature. By "improvement," Darwin meant the modification of traits toward configurations that benefit the individual in what he called the "Struggle for Existence," construed broadly as including "not only the life of the individual, but success in leaving progeny" (Darwin, 1859, p. 62). Darwin initially had no name for the abstract property being improved; he did not use the term *fitness*, which was coined later for this purpose based on Herbert Spencer's phrase "survival of the fittest," which Darwin adopted as a synonym for natural selection in later editions of the *Origin* (Gayon, 1998; Iseda, 1996). The classical usage of the term fitness is exemplified by statements that fitness "in the Darwinian sense" is assessed by "average number of offspring left" (Haldane, 1938, p. 78) and that "Darwinian fitness is measurable only in terms of reproductive proficiency" (Dobzhansky, 1962, p. 129). In other words, being fit is about creating descendants.

The crucial role of fitness in evolutionary theory has been well described by Brandon (2019):

> Why is it that some variants leave more offspring than others? In those cases we label natural selection, it is because those variants are better adapted, or are fitter than their competitors. Thus we can define natural selection as follows: Natural selection is differential reproduction due to differential fitness (or differential adaptedness) within a common selective environment […]. This definition makes the concept of natural selection dependent on that of fitness, which is unfortunate since many philosophers find the concept of fitness deeply mysterious (see e.g., Ariew & Lewontin, 2004). But like it or not, that is the way the theory is structured.

Social interactions are especially relevant in this context. In general, a gene is selected for if its phenotypic effects enhance the transmission of its identical copies, including copies in

Received October 14, 2023; revisions received February 21, 2024; accepted March 11, 2024
Associate Editor: Maria E. Orive; Handling Editor: Jason Wolf

other individuals. To account for this, Hamilton (1964) proposed that natural selection favors strategies that maximize an individual's inclusive fitness (henceforth, IF_{Hamilton}), defined as:

> the personal fitness which an individual actually expresses […] after it has been first stripped and then augmented in a certain way. *It is stripped of all components which can be considered as due to the individual's social environment, leaving the fitness which he would express if not exposed to any of the harms or benefits of that environment* [emphasis added]. This quantity is then augmented by certain fractions of the quantities of harm and benefit which the individual himself causes to the fitnesses of his neighbours. The fractions in question are simply the coefficients of relationship […] (Hamilton, 1964, p. 8).

According to Hamilton, a gene is selected for if it satisfies Hamilton's rule $rb - c > 0$, where r is relatedness, $-c$ and b are changes caused to the reproduction of "self" and "other" when a focal individual expresses the gene, and the expression $rb - c$ is called the gene's *inclusive fitness effect* (Grafen, 2006; Hamilton, 1964). A positive inclusive fitness effect implies that the gene's phenotypic effect increases the focal individual's IF_{Hamilton} as defined above.

Other definitions of fitness and inclusive fitness have also been proposed (Section 11). One recent proposal called the "folk definition of inclusive fitness" (henceforth, IF_{folk}; Fromhage & Jennions, 2019) differs from IF_{Hamilton} in omitting the highlighted part in Hamilton's quote above. Accordingly, IF_{folk} is the sum of an individual's own offspring (*including* any accrued due to the social environment) plus its effects on its relatives' number of offspring, weighted by relatedness.

Here, we ask what properties a fitness concept needs in order to capture natural selection's inherent tendency for cumulative improvement. We approach this question via a two-step procedure in which we first identify theoretical properties that motivate definitions of fitness and then ask how these properties complement each other to produce a workable theory of adaptive evolution. We then compare fitness concepts in light of these requirements. We find that, among the considered alternatives, only IF_{folk} fills the desired theoretical role under general conditions. To help readers navigate our arguments, we provide an outline of the main ideas (Box 1) and a nontechnical video (Supplementary Material).

Box 1. The main ideas in outline

(i) The concept of a "design principle"—which explains what evolutionary adaptations are for—is essential for thinking about long-term adaptation. Some other evolutionary questions, such as whether a particular gene will be selected for in the short term, do not require such a principle.

(ii) Currently, the most popular candidate for a design principle is that organisms evolve to maximize IF_{Hamilton}. However, IF_{Hamilton} cannot explain an important class of adaptations that involve nonadditive interactions between social partners (e.g., the synergistic cooperation between workers and royals in social insects). This is because IF_{Hamilton} strips away the reproductive success due to an individual's help from others.

(iii) The reason for this failure is that IF_{Hamilton} was intended to be both a design principle and a predictor of allele frequency change. Predicting adaptation is not the same as predicting short-term allele frequency change, however, because in the long term, the common interests of the "parliament of genes" will tend to win out over the idiosyncrasies of selection on particular genes (Hammerstein, 1996; Leigh, 1971).

(iv) Consequently, if our principal aim is to formalize a design principle, we should abandon the subsidiary goal of predicting short-term change in allele frequencies, and instead focus on the long-term outcome of adaptive evolution.

(v) IF_{folk} is a design principle that can explain long-term adaptation under general conditions, including cases with nonadditive interactions between social partners (such as in social insects).

(vi) Although it fails to predict allele frequency change in some cases, IF_{folk} can be reframed from a gene's-eye view, by focusing on a particular type of (Mendelian low penetrance, i.e., non-segregation distorting and rarely expressed) "reference gene."

2. How to think about adaptive evolution

A popular approach to understanding which traits are favored by natural selection is to substitute the question

(**2A**) Will trait T be selected for?

with

(**2B**) Will a gene inducing trait T be selected for?

This substitution of genes for traits is often done without worrying too much about the genetic details (e.g., Grafen, 1984) in the hope that such details will not matter enough to produce opposing answers to these two questions. This hope is only justified, however, if the envisaged gene is a suitable stand-in for what presumably is really going on in most cases of interest, namely cumulative multilocus evolution. Not every gene is suitable for this purpose: for example, almost any maladaptive trait (say, a preference for banging one's head against rocks) could spread under selection if encoded by a segregation-distorter gene. Although questions 2A and 2B may have different answers depending on the type of gene we are considering, we can envisage a class of genes for which these questions are equivalent. This raises the question:

(**2C**) What kind of gene should we envisage to make gene-level selection a good heuristic for organism-level adaptive evolution?

3. Two complications

A common way to set up population genetic models, usually without explicitly considering question (2C), is to envisage a Mendelian gene that is always expressed (i.e., that has high penetrance). In the context of social evolution, however,

the tendency of high-penetrance genes to be simultaneously expressed in interacting relatives causes two complications:

(**3A**) High-penetrance genes can cause social benefits to accrue disproportionally toward particular genotypes, even among potential recipients of the same pedigree relatedness to the focal individual. This can cause high-penetrance genes to be selected against even if the trait they code for is in the genome's majority interest. For example, imagine a situation in which helping a sibling provides large benefits but is exclusively directed to nonhelpers (Figure 1A). A full-penetrance *helping gene* is then selected against because of its absence in the beneficiaries of helping. This counterintuitive result has been called Charlesworth's paradox (McElreath & Boyd, 2007). Nevertheless, helping evolves in this situation under realistic genetic architectures that involve genes with any degree of penetrance (Figure 1B; Garcia-Costoya & Fromhage, 2021). We propose the term *social benefit distortion* for the biasing of social benefits toward particular genotypes by high-penetrance genes. Social benefits can also be distorted by other mechanisms, e.g., so-called greenbeard genes, which simultaneously encode a cue and a tendency to behave altruistically toward other cue-bearers (Dawkins, 1976). But because greenbeard genes are widely known to rely on genetic architectures that are often unrealistic and susceptible to invasion by alternative alleles (Gardner & West, 2010; Ridley & Grafen, 1981), they do not usually tempt us to view them as representative of cumulative multilocus evolution. Hence, we do not consider them further here.

(**3B**) High-penetrance genes make it difficult to separate causation and correlation: if a focal individual has a helping gene that causes it to generate benefits for its relatives, then this may correlate with this focal individual also receiving benefits. When calculating selection on the helping gene, one must, therefore, take care not to lump together the effects caused by a focal individual's action with effects that are merely correlated with this action. A failure to discriminate counts the benefit twice (once when provided and once when received) while counting the cost only once (Grafen, 1982; Levin et al., 2019). In the literature, this is known as the *double accounting problem*.

4. Hamilton's solution

Complications (3A) and (3B) can both be sidestepped by assuming that the total effect of multiple causes is simply the sum of their individual effects, which are independent of one another ("additive causality"; Birch, 2016). This assumption implies that the magnitude of the benefit from a given helping act does not depend on the recipient's phenotype (and hence genotype). This avoids complication (3A) by removing the source of bias (i.e., the *social benefit distortion*) toward particular genotypes. It also offers a simple way to handle

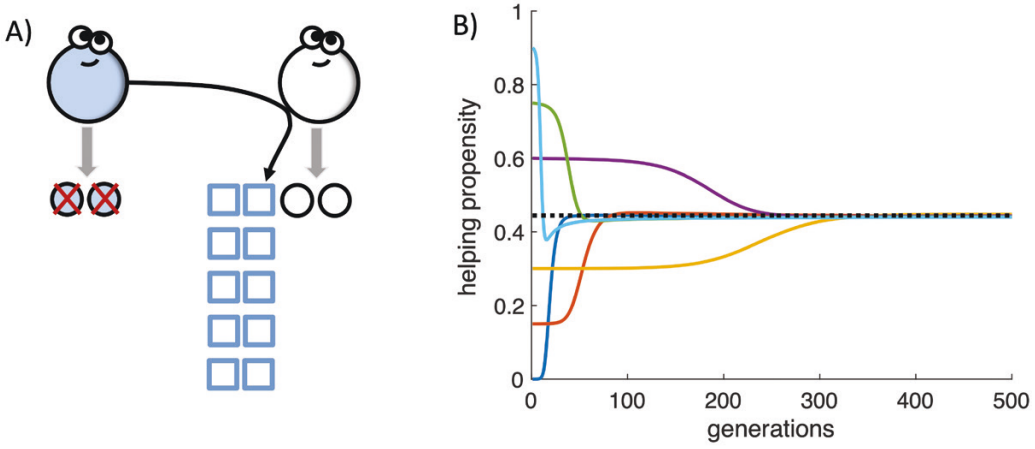

Figure 1. A case of gene-level selection for maladaptive organismal phenotypes driven by *social benefit distortion*. Suppose that in each generation, siblings face an opportunity to help each other, but help can only be directed toward nonhelpers. This introduces a nonadditivity, whereby a focal individual's received benefits depend on its own phenotype. (A) The individual on the left carries a helping gene, which induces it to help a recipient (related to the helper by $r = 0.5$). The helper sacrifices its own reproduction (crossed-out circles; cost $c = 2$) to make the recipient produce $b = 10$ extra offspring (the small squares) on top of its baseline production (the two small circles). Although helping in this scenario meets Hamilton's rule (since $10 * 0.5 > 2$), it is selected against *if encoded by a gene that is always expressed*. This outcome, called Charlesworth's paradox, arises because if every sibling with the helping gene expresses it, then the benefits of helping accrue only to those siblings who lack the gene. And, as far as the propagation of the focal gene is concerned, benefits that only accrue to siblings who lack the focal gene count for nothing. Superficially, Charlesworth's paradox appears to show that inclusive fitness theory rests on restrictive genetic assumptions (so-called "weak selection") that are not met in this example. By contrast, Garcia-Costoya and Fromhage (2021) have shown that the paradox does not arise under realistic genetic architectures that allow for gradual, cumulative change, which may involve various strengths of selection acting on various alleles. Under realistic conditions, helping evolves to the approximate level predicted based on a phenotypic interpretation of Hamilton's rule. With this result in mind, we propose the term *social benefit distortion* for the mechanism generating the biased flow of benefits toward siblings with particular genotypes that drives selection against helping in Charlesworth's paradox. Inspired by the familiar term *segregation distortion*, this new term highlights the similarity that both types of distortion can lead to the evolution of maladaptive phenotypes that face counter-selection from the so-called "parliament of genes" (Leigh, 1971). The possibility of gene-level selection for maladaptive organismal phenotypes implies that it is a mistake to think of selection-driven gene-frequency change as synonymous with adaptive evolution. (B) Simulated evolutionary trajectories when helping propensity (i.e., the probability that an individual will help) is encoded by allelic values between 0 and 1 at a haploid quantitative locus (analogous to Figure 1B of Garcia-Costoya & Fromhage, (2021)). Coloured lines correspond to different initial helping propensities. The dashed line shows the prediction based on individuals acting to maximize their IF_{folk}. Mutant variants are drawn from a uniform distribution, allowing steps of any size and strength of selection (i.e., not confined to "weak selection"). For details, see Supplementary Material.

complication (3B): since the effect of helping (in terms of additional offspring produced) is assumed not to depend on the receiving individual's phenotype, such additional offspring can be safely excluded ("stripped away") when asking what phenotype is favored by natural selection. This stripping procedure, which is built into the definition of IF_{Hamilton} (see Introduction), then correctly isolates a helping gene's causal effect from the correlational component mentioned in (3B).

5. The optimizing tendency of evolution and the purpose of life?

If the question "Will trait T be selected for?" (2A) could be answered in the form "a trait will be selected for if it improves the organism's phenotypic performance as judged by some measurable success criterion X," then identifying X would reveal what organismal property is targeted by the Darwinian tendency of cumulative improvement. In other words, identifying X would allow us to characterize the optimizing tendency inherent in evolution. Moreover, because a history of cumulative improvement should render organisms well-adapted to achieve high X (including, sometimes, behaviorally striving for X), we may metaphorically call X the "biological purpose of life." This emphasizes the connection between X and the apparently purposeful character of traits and behaviors shaped by natural selection.

6. A disappointment

Hamilton wrote about an individual's IF_{Hamilton} in ways which suggest that he envisaged it as the criterion X described in the previous section. Specifically, he showed that an allele is selected for if it increases the IF_{Hamilton} of its carriers and concluded from this that IF_{Hamilton} is a quantity which "each organism appears to be attempting to maximize" (Hamilton, 1964, p. 1).

However, IF_{Hamilton} sometimes strikingly fails to predict adaptive behavior because of its reliance on the unrealistic assumption of additive causality (see section 4). Consider the following variation on Haldane's (1955) famous thought experiment: Haldane stands on the shore, trying to decide whether he should jump into the water to save his children from drowning. From the standpoint of adaptive behavior, should he discriminate between child A produced without the help of the social environment and child B produced with such help? Presumably, if he were trying to maximize a quantity that excludes offspring produced with the help of the social environment (paraphrasing IF_{Hamilton}), then the answer should be "yes." But biologically, that is clearly the wrong answer because the children's prior history is irrelevant to their propagating Haldane's genes in the future. Nor is it reasonable to suppose that child B is necessarily beyond Haldane's control simply because the child's existence initially required help from others. The nonadditivity of this situation lies in the dependence of the value of one type of help on the existence of another type of help: if Haldane does not save child B, then the help that induced him to produce this child in the first place will have been in vain. Similar problems with IF_{Hamilton} have surfaced in other contexts, where they have been called Creel's paradox (see Queller, 1996; which is an important precursor to our present perspective as explained in Q23 of Fromhage & Jennions, 2019), Charlesworth's paradox (Figure 1), and Skyrms's paradox (Martens, 2019). Since there is nothing paradoxical about finding that a concept built on restrictive assumptions has limited applicability, this terminology gives a sense of the high hopes placed in IF_{Hamilton}.

7. Can gene-level considerations rescue IF_{Hamilton}'s generality?

To calculate selection at the level of genes, the restrictive assumption of additive causality is not needed (Gardner et al., 2011; Nowak et al., 2010; Queller, 1992). This is because, regardless of the causal underpinnings, once we know the reproductive success of all genotypes, we can use this information to calculate the gene-frequency change from one generation to the next. Since such calculations can produce generalized forms of Hamilton's rule (i.e., forms that do not assume additive causality; see Birch & Okasha, 2015 for a review), one might think that, by extension, they rescue the general applicability of IF_{Hamilton}. For example, a group of 137 authors (Abbot et al., 2011), responding to a controversial article by Nowak et al. (2010), claimed that, according to "the completely general theory of inclusive fitness," "natural selection leads organisms to become adapted as if to maximize inclusive fitness." This claim is unjustified, however, because predicting adaptation is not the same as predicting short-term allele frequency change. (Recall the nonequivalence of questions 2A and 2B about phenotypic vs. gene-level selection.) Hence, gene-level calculations do not vindicate IF_{Hamilton} as a general goal of organism-level adaptive behavior.

There are different senses in which models can be said to be "additive." As explained in section 4, the use of IF_{Hamilton} can be justified if it is valid to assume that the effects of multiple traits, e.g., behaviors, combine by adding up (additive causality). This implies that offspring produced through help from others are beyond the focal individual's control. Because these offspring will arise no matter what the focal individual does, they become irrelevant to the question of what the focal individual should do so as to best propagate its genes. Additive causality is a strong ecological assumption about how traits interact, including traits encoded by different loci. (It is hard to envisage a realistic situation in which, for example, selection could not act on a breeder to change the extent to which it converts extra provisioning by helpers into more offspring.) Additive causality is not to be confused with the much weaker assumption that the effects of trait deviations *caused by a given allele* combine additively. The latter assumption is approximately true for a large and important class of alleles (namely those inducing small trait deviations, so-called δ -weak selection; Wild & Traulsen, 2007; and those slightly changing the probability that a trait is expressed, so-called *probabilistic mixing* Levin & Grafen, 2021). Gene-level additivity is a popular simplifying assumption (e.g., Scott & Wild, 2023; Taylor & Frank, 1996) which, we believe, has a good biological justification in that it avoids social benefit distortion, hence the evolution of maladaptive phenotypes (section 3A; Figure 1). But gene-level additivity does not place any offspring beyond the focal individual's control, so it does not rescue IF_{Hamilton}'s general applicability. For example, it does not resolve the drowning scenario of section 6, whereby IF_{Hamilton} predicts that organisms should not save offspring that were produced with the help of others.

8. A fresh start

In view of the difficulties with IF_{Hamilton}, let us go back a few steps to approach the complications of section 3 (social benefit

distortion and the confounding of causation and correlation) from a different angle. Instead of a high-penetrance gene, can we identify a better type of gene to act as a heuristic for cumulative, multilocus evolution? Envisage a low-penetrance (i.e., rarely expressed) gene that is expressed by a focal organism but not by other individuals in its immediate social environment. Then adopt a *gene's-eye view* (Dawkins, 1976) and ask: what phenotype should this gene induce to maximize the transmission of identical copies? We can immediately note that answering this question is made easier because problems 3A and 3B do not arise for a gene with very low penetrance. (Hence, there is no need to assume additive causality to solve them—see section 4.) Because a low-penetrance gene is almost never simultaneously expressed in interacting individuals, any nonadditivities that might arise in that rare event will have a negligible effect on selection. Hence, a low-penetrance gene has approximately additive effects, like the genes involved in δ-weak selection and probabilistic mixing described in the previous section. This avoids the problem of social benefit distortion (see section 3A), which, in turn, ensures that the optimal phenotype for this gene will also aid the transmission of most other genes contained in the same organism. This property makes a low-penetrance gene a suitable stand-in for the genome's "majority interest," which is to shape organisms that function well as coherent wholes (Leigh, 1971; Patten et al., 2023). Accordingly, Fromhage & Jennions (2019) proposed a Mendelian (i.e., not segregation distorting), low-penetrance gene as the answer to question 2C. They refer to such a gene as a *reference gene*.

Adding up a reference gene's transmitted copies (specifically, the copies causally attributable to a focal individual) gives a quantity proportional to that individual's IF_{folk} (Fromhage & Jennions, 2019; section 2). Hence, an individual's IF_{folk} is proportional to the transmission success of a reference gene carried by this individual. As we explain in the *Discussion* section, this does not necessarily mean that a gene will be positively selected if it has a positive *correlation* with IF_{folk} across individuals in a population. However, if the reference gene idea works as intended in delineating what phenotypic changes *caused* by genes conform to the genome's majority interest, then cumulative changes throughout the genome should tend to shape organisms toward IF_{folk}-enhancing phenotypes. If true, this would avoid the so-called paradoxes that have stymied IF_{Hamilton}'s use in predicting adaptive phenotypes (see section 6). On this view, sweeping claims of equivalence between the gene's-eye view and inclusive fitness are misleading because gene-level selection can favor traits that decrease their bearer's inclusive fitness (e.g., due to social benefit distortion—see Figure 1). Nevertheless, to the extent that the parliament of genes vetoes maladaptive (i.e., inclusive fitness-decreasing) traits, phenotypic change will optimize inclusive fitness in the long run. There are four main lines of argument that support this view.

(**8A**) Fromhage and Jennions (2019, Q14) prove that, at an evolutionary equilibrium, individuals must express a phenotype that maximizes their IF_{folk}. In short, their argument runs like this: Because IF_{folk} is proportional to a reference gene's propagation success, the statement "the mutant gene increases the IF_{folk} value of an individual expressing it" is equivalent to "the mutant gene increases its transmitted number of copies." Thus, unless individuals already express the IF_{folk}-maximizing phenotype, there is scope for IF_{folk}-increasing low-penetrance mutations to be selected for. This proof assumes only that low-penetrance mutations arise, not that they are common.

(**8B**) Because an individual's IF_{folk} is proportional to the transmission success of a reference gene carried by this individual, any focal gene's causal effect on the individual's IF_{folk} predicts the direction of selection on other genes that would modify the focal gene's expression. For example, if expressing (compared to silencing) a focal gene decreases a focal individual's IF_{folk}, then it follows that a modifier gene which suppresses the focal gene's expression would increase IF_{folk}. This implies positive selection for the modifier gene (for a formal example, see Q15 in Fromhage & Jennions, 2019), at least if it meets the definition of a reference gene. (The latter proviso rules out the possibility of social benefit distortion. For example, one could construct a version of Charlesworth's paradox (Figure 1) in which a full-penetrance modifier gene suppresses the nonhelping allele's expression.) To grasp the biological relevance of this argument, it is important to realize that high- vs. low-penetrance genes do not have conflicting evolutionary interests (Queller, 2019). Instead, in situations where high- vs. low-penetrance genes drive phenotypic evolution in opposite directions, we can think of high-penetrance genes as clunky building blocks that "overshoot the target" and cannot build anything useful, except in combination with smaller blocks. For instance, when it is optimal to express trait T with intermediate probability, then a high-penetrance gene encoding T cannot implement the optimal strategy on its own. It can do so, however, if its expression is suitably modified by other genes.

(**8C**) Since the true genetic complexity of adaptive evolution defies full mathematical description, we need to know what simplifying assumptions best capture the essential features of adaptive evolution. This, however, requires an independent source of information about what these essential features are. Simulation studies support the idea that adaptive evolution tends toward IF_{folk}-enhancing phenotypes, leading to approximate maximization at equilibrium under realistic genetic architectures that allow (but do not enforce) gradual, cumulative change (Garcia-Costoya & Fromhage, 2021; SM2 in Fromhage & Jennions, 2019). This gradual change need not be primarily driven by low-penetrance genes. Low-penetrance genes, being weakly selected, are a minor evolutionary force. But that does not preclude phenotypic change from being largely driven by genes (with various degrees of penetrance) whose phenotypic effects are qualitatively in line with the genome's majority interest (see also Q1 and Q3 in Fromhage & Jennions, 2019).

(**8D**) Under suitable simplifying assumptions (in particular so-called δ-weak selection, which focuses on genes with small, approximately additive effects; Grafen, 1985; Wild & Traulsen, 2007), evolution follows a marginal version of Hamilton's rule $rb_m - c_m > 0$, in which r is pedigree relatedness and $-c_m$ and b_m can be interpreted as causal effects which a marginal increment in a focal individual's trait value has on reproductive success (of the focal individual and its relative, respectively; Taylor et al., 2007). According to this rule, the criterion for a phenotypic change to be positively selected is that it has a positive total effect on the direct and indirect fitness components that make up IF_{folk}. This predicts gradual evolution toward IF_{folk}-enhancing trait values, with local maximization at equilibrium. By contrast, Hamilton's marginal rule does not generally predict evolution toward IF_{Hamilton}-enhancing traits. This is because offspring produced with the help of the social environment are excluded from IF_{Hamilton}, whereas small changes in the number of such offspring are included in Hamilton's marginal rule.

9. A wish list for fitness enthusiasts

We identify five theoretical properties, corresponding to five roles played in evolutionary reasoning, which may motivate calling a variable *fitness*:

(**9A**) Predictor of (short-term) phenotypic change (see question 2A).

(**9B**) Predictor of gene-frequency change (see question 2B).

(**9C**) Improvement criterion inherent in natural selection (see section 5).

(**9D**) Performance measure for phenotypic strategies.

(**9E**) Performance measure for individual organisms.

The distinctness of properties 9A and 9B follows directly from the nonequivalence of questions about phenotypic vs. gene-level selection (2A and 2B). The distinction between 9D and 9E follows from the observation that, since each individual lives only once, no individual can be cumulatively improved over the generations. What can be cumulatively improved, instead, is the strategy encoded by the whole genome, which prescribes what phenotype to exhibit in given circumstances. Because an improvement criterion (9C) may be limited to comparing specific alternatives (e.g., phenotypes that differ in one trait at a time), unlike 9D and 9E, it does not capture the idea that numerous traits must be coadapted to each other to produce strategies and organisms that perform well as coherent wholes. 9D and 9E are closely linked to the notion of *adaptedness*, which refers to an entity's propensity to perform well in a given environment. Here, by "to perform," we mean "to exhibit traits and behaviors that are conducive to success." Adaptedness and performance should be measured in terms of *expected* success because realized success is partly a matter of luck. Even the best-adapted individual may fail to reproduce if struck by lightning (Brandon, 1978; Brandon & Beatty, 1984; Mills & Beatty, 1979), and even the best-performing individual may fail to leave descendants if its offspring are struck by lightning. Nevertheless, in practice, propensities are estimated by averaging across observed outcomes, so biologists commonly use the term *fitness* to refer to both propensities and outcomes.

10. The logic of adaptive evolution

Once we know what traits are favored by natural selection, the basic idea of adaptive evolution is simple. For example, if height-enhancing genes are consistently selected for, then, other things being equal, the accumulation of such genes over evolutionary time will cumulatively increase height. Let us now consider what properties a fitness concept should have to make this idea work in general.

(**10A**) If a fitness concept could be found that has properties 9A–E, the logic of adaptive evolution could be summarized as follows: any gene that improves the phenotypic strategy (see 9D), as measured by its carriers' performance (see 9E), is favored by natural selection (see 9B), thus directing short-term phenotypic change (see 9A) in a way that amounts to cumulative improvement (see 9C) in the long run. Sadly, this simple scheme fails because no fitness concept has yet been proposed (and we doubt one exists; see Figure 1) that combines all five properties under general conditions.

(**10B**) If a fitness concept could be found that has properties 9C–E, and that additionally has properties 9A–B in a heuristic sense, the logic of adaptive evolution could still work along the same lines: genes that improve the phenotypic strategy (see 9D) as measured by their carriers' performance (see 9E), are *usually* favored by natural selection (heuristic use of 9A and 9B), creating trends of cumulative improvement (see 9C) despite the potential for short-term change in other directions.

(**10C**) For comparison, we also sketch an (unsuccessful) attempt to summarize the logic of adaptive evolution with a gene-centric fitness concept that only has properties 9A and 9B. By giving the name "fitness" to some metric that indicates whether a gene is positively selected, we can truthfully (but trivially) say that genes with higher fitness are consistently selected for. From there, it appears a small step to claim that genes that confer high fitness to *organisms* are consistently selected for, driving cumulative change toward better-adapted organisms. This latter step is unjustified, however, because it rests on an illegitimate shift from a gene-centric to an organism-centric meaning of "fitness."

To emphasize the special status of any fitness concept that encapsulates the logic of adaptive evolution according to either scheme 10A or 10B, we will characterize such a concept as a *design principle* of adaptive evolution (an expression coined for similar purposes by West & Gardner, 2013). In the metaphorical language of section 5, a design principle encompasses both the optimizing tendency of evolution and the biological purpose of life. By contrast, scheme 10C reveals very little about the adaptive evolution of organisms.

11. Comparing fitness concepts on theoretical grounds

A comparison of fitness concepts and similar constructs (Table 1) reveals three candidates which qualify as design principles, at least under some conditions: IF_{Hamilton}, IF_{folk}, and lifetime reproductive success (*LRS*). IF_{Hamilton} has properties 9A–E, thus conforming to scheme 10A, but only under the restrictive conditions where it can be applied (in particular, additive causality). IF_{folk} has properties 9C–E and, in a heuristic sense, 9A–B, thus conforming to scheme 10B under broad conditions that include nonadditive causality. *LRS* is a special case of both IF_{Hamilton} and IF_{folk}, which assumes the absence of social interactions between relatives. A technical caveat about uniqueness is needed here. In game theory, a utility function is only ever unique up to the choice of origin and unit, implying the existence of a family of transformations that are all maximized by the same strategy (Okasha & Martens, 2016). In other words, there is no practical difference between striving to maximize U or $a + b^*U$, where a is any constant and b is any positive constant, because the U-maximizing behavior will necessarily maximize $a + b^*U$ as well. Likewise, each of the design principles mentioned above has arbitrarily many transformations that would make the same empirical predictions and could properly be considered to represent the same design principle. Not all transformations, however, lend themselves equally to biological interpretation. For fitness to measure Darwinian performance, it seems reasonable to demand that it be zero for an organism that neither reproduces nor helps any relatives. This outcome then sets a biologically meaningful baseline against which to measure Darwinian performance.

12. Comparing fitness concepts on empirical grounds

Let us now assess our candidate design principles in light of empirical examples. Figure 2 shows a selection of organisms

Table 1. Comparison of the theoretical properties of different fitness measures.

	Properties	Assumptions	Remarks
Lifetime reproductive success (*LRS*)	9A–E	Mendelian inheritance, no social effects	For the special case of nonoverlapping generations, the Price equation (Price, 1970) shows that *LRS* predicts gene-frequency change (9B), implying short-term phenotypic change (9A) to the extent that genes have additive genetic effects. This allows cumulative improvement (9C) of any traits that consistently promote *LRS*, which in turn justifies interpreting *LRS* as a criterion for a strategy's performance (property 9D) as measured by the performance of the individuals adopting it (property 9E). More generally, an evolutionarily meaningful measure of *LRS* must weigh offspring by their reproductive value, i.e., by each offspring's expected contribution to the future gene pool (Fisher, 1930; Williams, 1966b). For example, if instead of producing one typical offspring, a parent could produce X offspring whose per-capita contribution to the future gene pool was Y times higher (such that $X*Y > 1$; e.g., due to improved survival and/or reproduction), then by choosing the latter option the parent will increase its own contribution to the future gene pool. Grafen (1999) showed that similar considerations can take care of changes in population size: because one individual among few makes a larger proportional contribution than one among many, each offspring's reproductive value is inversely related to population size at its time of birth. Thus, while the timing of reproduction matters insofar as it places offspring in different contexts, natural selection nevertheless favors traits that increase (reproductive value-weighted) *LRS* (Grafen, 1999). For analogous reasons, weighting offspring by their reproductive value is also needed in all other offspring-based fitness measures. Omitting this complication—as we do below for simplicity—can be justified by assuming that all offspring have equal reproductive value.
IF_{Hamilton}	9A–E	Mendelian inheritance, weak selection, additive causality	For a definition, see *Introduction* section. Hamilton (1964) showed that genes positively correlated with IF_{Hamilton} are selected for, establishing property 9B. Because Hamilton's "stripping procedure" excludes the noncausal component mentioned in section 3B, this is equivalent to saying that genes which causally increase IF_{Hamilton} are selected for. This implies selection for IF_{Hamilton}-enhancing traits, establishing property 9A. Moreover, Hamilton showed that if an equilibrium exists, then the strategy exhibited at equilibrium should maximize the expected IF_{Hamilton} of the individuals adopting it (also see Grafen, 2006). These results justify interpreting IF_{Hamilton} as a criterion for a strategy's performance (property 9D) as measured by the performance of the individuals adopting it (property 9E). Trait changes toward better performance can then be interpreted as improvements (property 9C). The assumption of weak selection is needed to ensure that the "pedigree relatedness" invoked in the definition of IF_{Hamilton} accurately reflects relatedness with respect to the focal gene (Grafen, 1985). However, without the assumption of additive causality, IF_{Hamilton} does not qualify as a quantity which organisms should appear adapted to maximize (see sections 6 and 7; also see section 7 in Fromhage & Jennions, 2019).
IF_{folk}	9C–E	Mendelian inheritance, low-penetrance mutations occur at the positive frequency	For a definition, see *Introduction* section. According to the logical proof mentioned in section 8A, at an evolutionary equilibrium, individuals must express a phenotype that maximizes their IF_{folk}. According to the arguments presented in sections 8B–D, long-term evolution should tend toward IF_{folk}-enhancing phenotypes. These arguments also justify interpreting IF_{folk} as a criterion for a strategy's performance (property 9D) as measured by the performance of the individuals adopting it (property 9E). Trait changes toward better performance can then be interpreted as improvements (property 9C). Additionally, IF_{folk} has properties 9A and 9B in a heuristic sense, namely for those gene-frequency and phenotypic changes that conform to the predicted long-term trend. In this context, properties 9A and 9B are to be interpreted causally, such that alleles (or traits) which cause their individual bearer's IF_{folk} to increase are heuristically predicted to be selected for (as in Figure 1 of Fromhage & Jennions, 2019; but see their Q15 for a counterexample). The relevant notion of performance (9E) quantifies an organism's overall causal effect as determined by the entirety of its developmental and behavioral choices from conception onwards (for further details, see Q17 in Fromhage & Jennions, 2019). As a performance measure, IF_{folk} quantifies what an individual makes of its circumstances without implying that outcomes achieved under different circumstances are directly comparable—see discussion about confounders in *Discussion* section. This qualifier is needed because high-penetrance genes may simultaneously modify a focal individual and its social environment (see section 3B).
IF_{folk} + (arbitrary constant)	9C–D	Mendelian inheritance, low-penetrance mutations occur with non-zero frequency	Adding a constant undermines the interpretation as a performance measure (i.e., property 9E). To see this, consider the following analogy. Take a person's monetary earnings as a measure of their economic performance. Then, adding to this quantity an arbitrary constant (say, the person's birthweight in grams converted to €) yields a composite quantity that no longer justifies this interpretation. This is true even though the composite quantity is still equally useful for decision-making based on comparing its predicted values under possible courses of action. In the context of fitness, the added arbitrary constant might refer to that part of the reproduction of relatives which is not causally affected by the focal individual. The corresponding composite quantity is called "simple weighted sum" including fitness (Grafen, 1982; also see Q26 in Fromhage & Jennions, 2019).
Neighbor-modulated fitness (*NMF*)	9A–B	Mendelian inheritance	An individual's reproductive success is modulated by social interactions with its neighbors (Hamilton, 1964). Because performing a costly helping act (compared to not performing it) by definition decreases the actor's reproductive success (see SM4 in Fromhage & Jennions, 2019), an individual performing such acts will not thereby maximize its *NMF*. Hence, *NMF* captures neither an individual's (9E) nor a phenotypic strategy's (9D) performance, which could be invoked as an improvement criterion (9C).

Table 1. Continued

	Properties	Assumptions	Remarks
Hamilton's marginal rule (sensu Birch & Okasha, 2015)	9A–B	Mendelian inheritance, δ-weak selection (see section 8D)	Has been argued to have also property 9C (Birch, 2017a, 2019), but without identifying a corresponding quantity which could have properties 9D and 9E. To see why this is problematic, consider the following analogy. If we had a crystal ball that recommends changes to our economic policy, the only way to establish that these changes qualify as improvements would be to identify a quantity that is thereby being improved—say, our bank account balance. Likewise, attributing property 9C to Hamilton's marginal rule arguably requires linking it to a quantity that would be improved by trait changes satisfying the rule. In section 8D, we argue that this quantity is IF_{folk}.
Hamilton's general rule (sensu Birch & Okasha, 2015)	9A–B	Mendelian inheritance	A version of Hamilton's rule $r_g b_g - c_g > 0$ in which b_g and c_g are partial regression coefficients fitted to decompose a known evolutionary change (Queller, 1992), and r_g is a measure of trait-specific genetic similarity ("regression relatedness" Hamilton, 1972; not to be confused with pedigree relatedness). Arguably, this rule has properties 9A and 9B, although some authors deny that it predicts anything (Allen et al., 2013; van Veelen et al., 2017). The issue is that because each individual's genotype and reproductive success enter the calculation of b_g and c_g, in some sense the relevant evolutionary change must already be known in order to be "predicted" based on b_g and c_g.
Inclusive fitness sensu Akcay and van Cleve (2016)	9A–B	Mendelian inheritance	A measure of an allelic lineage's success is calculated from the average reproductive success of the relevant individuals. Ågren (2021) commented that "many of the potential pitfalls of inclusive fitness theory can be avoided if inclusive fitness is re-conceived as a property of the genetic lineage rather than of the individual organism." This is true, but it raises the question of why such a loaded term should be used for what is essentially a selection coefficient.
Inclusive fitness sensu Lehmann and Rousset's (2020) eq. 4/ eq. C.28	9A–B	Mendelian inheritance	A sum of direct and indirect fitness components. While their direct fitness component implicitly includes all of an individual's own offspring (i.e., there is no "stripping" as in IF_{Hamilton}), their indirect fitness component includes only the marginal effect "stemming from a single […] gene copy switching to expressing a copy of the mutant instead of the resident allele" (Lehmann & Rousset, 2020, p. 723). In other words, among a focal individual's effects on its relatives, all those effects *not* stemming from a particular mutant gene (but rather from established adaptations) are excluded. This makes Lehmann and Rousset's *IF* neither an absolute nor a marginal measure of reproductive success, but a mixture of both: it is absolute with regard to direct fitness, yet marginal with regard to indirect fitness. This partly marginal quantity is difficult to interpret biologically but certainly lacks property 9E. It implies, for example, that a sterile worker ant's *IF* is approximately zero (give or take the effect of a mutant allele) and that a worker bee's *IF* approximates her direct fitness from laying the occasional unfertilized egg, despite her phenotypic design and behavior being overwhelmingly dedicated to obtaining indirect fitness.
Inclusive fitness sensu Levin and Grafen (2021)	9A–B	Mendelian inheritance, weak selection	The sum of three components: "baseline asocial fitness, the difference to personal fitness as a result of the strategy, and relatedness weighted difference to social partners' fitnesses as a result of the strategy" (Ibid., p. 5). Here, the phrase "as a result of the strategy" refers to the difference arising from expressing the mutant strategy instead of the "incumbent strategy" played by almost all other individuals in the population. Like Lehmann & Rousset's *IF*, this is a partly marginal formulation that excludes effects of established adaptations (i.e., those which mutant and incumbent strategies have in common) and so does not capture an organism's overall success. Apart from its "asocial" component, Levin & Grafen's *IF* boils down to a criterion for the pairwise comparison of phenotypes, which lacks properties 9D–9E and arguably also 9C (see the remarks above on Hamilton's marginal rule).
Fitness according to Fisher's fundamental theorem (Fisher, 1930)	9A–B, 9C in a limited sense	Mendelian inheritance	A quantity which, when averaged across individuals, yields the population growth rate (Price, 1972). Designed to have property 9C but succeeds only in a limited sense. Based on the *correlation* between genotype and reproductive success, it essentially predicts selection for genes that have the good fortune of finding themselves in the bodies of successful reproducers. Says nothing about what causal properties of organisms are selected for. Although Fisher may have intended his theorem to be causal (Lee & Chow, 2013), his treatment neither accounts for, nor excludes by assumption, the possibility that the correlation may arise through social effects (Price, 1972).
Fitness according to Grafen (2015, 2018, 2019, 2020)	9A–B, 9C in a limited sense	Mendelian inheritance	Based on his generalization of Fisher's Fundamental Theorem to age-structured populations, Grafen (2015, p. 8) states that, rather than being "a tombstone evaluation of an individual's Darwinian success," fitness is defined at every moment in time such that "the mean value of fitness is the same in each age class, and equals the Malthusian parameter." The motivation behind this definition is that to ensure the truth of the statement that (the additive genetic component of) "average fitness" (as measured by reproductive-value-weighted contributions of all individuals) always increases, age-specific fitnesses must be normalized in a certain way. This normalization makes fitness relative to the average individual of the same age class. Like Fisher's original, this version of fitness says nothing about organisms' causal properties. Moreover, its focus on a given moment in time neglects the fact that phenotypic strategies succeed or fail as integrated wholes, depending on whether traits expressed at different times are suitably attuned to each other. For example, it is futile to ask whether building up fat reserves in autumn is good or bad *per se* (i.e., whether it increases fitness) because the answer depends on how the individual will spend the winter (Houston et al., 2023).

Table 1. Continued

	Properties	Assumptions	Remarks
Fitness as a phenotypic strategy's growth rate (e.g., Mylius & Diekmann, 1995)	9A–D	Clonal reproduction	A strategy's growth rate is not a property of an individual and cannot measure an individual's performance (9E). Under clonal reproduction, a variant of schema 10A applies, with performance evaluated at the level of the clonal lineage: any gene that improves the phenotypic strategy (see 9D), as measured by the lineage's expected growth rate, is favored by natural selection (see 9B), thus directing short-term phenotypic change (see 9A) in a way that amounts to cumulative improvement (see 9C) in the long run. By contrast, in sexually reproducing populations, where every individual is genotypically and phenotypically unique, there may be no discrete types whose growth rates could be cumulatively improved. And where discrete types do exist (e.g., color morphs), they are not the kinds of entities whose growth rates tend to be systematically improved through multilocus evolution. For example, although altruism among unrelated members of the same color morph might increase the morph's growth rate, this does not imply that such behavior is selected for. A possible source of confusion is the trivial fact that any morph that becomes predominant, for whatever reason, can be retrospectively assigned a high relative growth rate. Yet this neither explains why phenotypic change occurs (Ariew & Lewontin, 2004; Lewontin, 1983) nor implies that the change increases adaptedness.
Invasion fitness (e.g., Geritz et al., 1998)	9A–B	Clonal reproduction, or Mendelian inheritance in genetically almost uniform population	Refers to the growth rate of either a rare asexual strain or of a rare Mendelian allele in an otherwise genetically uniform sexual population. Predicts what phenotypic traits with such genetic underpinnings would be favored by selection. Does not measure an individual's performance (9E) nor engage directly with the ideas of phenotypic improvement (9C) or adaptedness. However, see the preceding fitness concept for a sense in which an asexual strain's growth rate may have properties 9A–D.
Fitness sensu McGraw and Caswell (1996)	–	N/A	A population growth rate is estimated from an individual-specific projection matrix. Lacks properties 9A–E because it is logically incoherent. Aimed to address the concern that unweighted *LRS* is not always an appropriate fitness measure. While it is true that, say, early produced offspring disproportionally increase their parent's contribution to a growing population's gene pool (compared to later-produced offspring, who miss out on contributing to this growth in the intervening time), the same is not true in demographically stable populations, where reproductive timing does *not* matter (Mylius & Diekmann, 1995). So, the demographic context is crucial for how different offspring should be weighted (see the remarks above on *LRS*). By contrast, McGraw and Caswell's approach gives different weights to early vs. late-produced offspring regardless of demographic context—as if each individual existed in a population of its own.

Note. The listed assumptions are needed to justify five key properties, namely 9A: Predictor of (short-term) phenotypic change; 9B: predictor of gene-frequency change; 9C: improvement criterion relevant to natural selection; 9D: performance measure for phenotypic strategies; 9E: performance measure for individual organisms.

Figure 2. (A) Bristlecone pine. (B) Bird of paradise. (C) Worker honeypot ant. (D) Termite queen. Photograph credits: (A) Rick Goldwaser from Flagstaff, AZ, USA—Gnarly. (B) Serhanoksay, CC BY-SA 3.0. (C) Greg Hume at en.wikipedia, CC BY 2.5. (D) 123rf.com stock photograph.

which, judged by their morphology, ecology, and behavior, appear adapted for strikingly different ends. The bristlecone pine (Figure 2A) reaches over 5,000 years of age. The male bird of paradise (Figure 2B) is attractive to females but at the cost of being conspicuous to predators. The honeypot ant worker (Figure 2C) is reproductively sterile and serves as a food storage container for its colony. The termite queen (Figure 2D) lays thousands of eggs daily but cannot feed herself. If all these organisms have been shaped according to the same design principle, then this fact should be reflected in their adaptation to pursue the same abstract life goal. This raises the question: what life goal do they all share? Recall our candidate design principles from section 11. Have all these organisms been adapted to have high *LRS*? No, the honeypot ant worker never reproduces. Are they adapted to spread their genes by means that don't rely on help from the social environment (paraphrasing IF_{Hamilton})? No, the termite queen is entirely dependent on help from others. Are they adapted to spread their genes by reproducing and/or helping relatives (paraphrasing IF_{folk})? Yes. This seems highly plausible, leaving IF_{folk} as our best candidate for the general design principle of adaptive evolution.

13. Discussion

The nature of biological adaptation has long been debated. Some philosophers have sought to illuminate the issue with refined fitness concepts (e.g., Brandon, 1978; Brandon & Beatty, 1984; Mills & Beatty, 1979; Sober, 1984), but others have dismissed the whole idea of adaptedness as metaphysical baggage inherited from natural theology (Byerly & Michod, 1991; Krimbas, 2004). More recently, after discussing IF_{Hamilton}'s limitations, Okasha (2018, p. 116) concluded that "there is no theoretical principle to the effect that natural selection will tend to produce adaptation, contrary to what is often thought." Likewise, he argued that the organism-as-rational-agent analogy (roughly, the idea that there is a link between the outcome of evolution and a goal that organisms behave as if they are trying to achieve; Grafen, 1999) lacks justification for want of a suitable goal.

Undeterred by these negative conclusions, here we make a new attempt to justify a general design principle of adaptive evolution. We argue (a) that genes which improve a phenotypic strategy, as measured by the expected IF_{folk} value of the individuals adopting it, tend to be favored by natural selection; (b) that this enables trends of cumulative improvement toward IF_{folk}-enhancing traits; and (c) that this, in turn, shapes adaptations that make organisms appear to be trying to maximize their IF_{folk}.

What good is a design principle?

Since one can model evolution without invoking a design principle (Allen & Nowak, 2016; Allen et al., 2013; Byerly & Michod, 1991; Doebeli et al., 2017), it is worth spelling out what is at stake. What do we gain from having a general evolutionary design principle?

First and foremost, such a principle allows us to explain the complexity, coherence, and adaptedness of organisms by the Darwinian mechanism of cumulative improvement toward better-adapted phenotypes. For adaptive evolution to climb "Mount Improbable" (Dawkins, 1996), the notion of "upward" in this metaphor must have a counterpart in reality. Indeed, in view of this explanatory need, even recognizing an imperfect design principle (such as *LRS*) seems a better theoretical choice than recognizing none.

Second, the design principle predicts what organisms should be like. It allows us to distinguish phenotypes that may plausibly evolve by natural selection from other imaginable phenotypes that will not. In the words of Gardner (2009), "It is possible to imagine worlds where organisms are designed to boil water or write poetry. The empirical value of

Darwinism owes to it identifying that, in this world, organisms are designed to achieve [direct and indirect] reproductive success." This entails testable predictions about how organisms should behave: provided an individual is well-adapted to its present circumstances, it should tend to behave in ways that increase its fitness.

Third, the design principle allows us to distinguish between adaptive and nonadaptive traits. When a warbler feeds a cuckoo chick (e.g., Davies, 2011), we can judge this behavior as nonadaptive if it decreases the warbler's fitness, even though the specific movements and routines involved are a product of natural selection (in another context). This judgement deepens our understanding of the situation, e.g., with regard to how selection would act on new variation in, say, warblers' cuckoo-detection capacity.

Fourth, it justifies how biologists talk about their subject matter. Many biological statements rest on the implicit assumption that there is a meaningful standard by which to judge the proper functioning of organisms and their traits (Haig, 2020). Examples range from "the hairs about the eye-lids are for the safeguard of the sight" (Bacon, 1623, 120) to "in order to gain access to buried stretches of DNA inside nucleosomes, a chromatin remodeling ATPase is required to unwrap the nucleosomal DNA" (Mellor, 2005, 147). Such statements are justified insofar as they can be grounded in a design principle as outlined by Neander (1991):

> It is the/a proper function of an item (X) of an organism (O) to do that which items of X's type did *to contribute to the inclusive fitness* [emphasis added] of O's ancestors, and which caused the genotype, of which X is the phenotypic expression, to be selected by natural selection.

Fifth, it allows us to draw a principled distinction between adaptive evolution and other evolutionary processes. Only adaptive evolution tends to systematically increase the fit between organism and environment in the relevant sense, namely, toward traits which confer high individual-level fitness in that environment (Fromhage & Houston, 2022; Welch, 2023; Williams, 1992).

Sixth, it allows us to see why evolutionary models need not make realistic genetic assumptions to produce biologically meaningful results, so long as they capture the relevant design principle. In particular, this accounts for the commonly made assumption that phenotypic evolution can be modeled as if unconstrained by genetic architecture (the so-called "phenotypic gambit"; Grafen, 1984).

Since much of biology is conducted in ways that take one or more of the above points for granted, it seems highly desirable to set these practices on a solid foundation. Indeed, Grafen (2018) goes so far as to call (his preferred justification for) characterizing natural selection as an improving process as "the air breathed by whole-organism biologists."

Predictive strengths and limitations

Because of the complication mentioned in section 3A, IF_{folk} has limited predictive accuracy about short-term evolutionary change. For example, in Charlesworth's paradox (Figure 1A) a gene for helping can be selected against despite helping behavior being highly adaptive (i.e., IF_{folk}-enhancing) at the phenotypic level. Is this predictive limitation a fatal problem? For the purposes of empirical biology, we think that the answer is "no." It is well understood that purely phenotypic predictions about evolution are only heuristic (e.g., Hammerstein, 1996; Marrow et al., 1996). For the purposes of mathematical modeling, however, this predictive limitation is certainly inconvenient. For example, since one cannot mathematically prove something that is not always true, there is no point trying to prove in general for IF_{folk} what Grafen (2006) proved (given additive causality) for IF_{Hamilton}: namely, that "natural selection *always* changes gene frequencies in the direction of increasing inclusive fitness" (Ibid., p. 543). More generally, if we accept that adaptive evolution can part ways with selection on high-penetrance genes (see 3A), this should, perhaps, reduce our trust in extrapolating from simple population genetic models to biological reality. But nature is under no obligation to make life easy for modelers, so these inconveniences are irrelevant to whether or not IF_{folk} correctly captures the design principle of adaptive evolution. Indeed, phenotypic models may predict long-term evolution better than genetic models because genetic details are liable to change in the long run (Hammerstein, 1996; Marrow et al., 1996).

Performance measure or correlational predictor?

Gene-frequency change under natural selection can be mathematically described using the covariance between genotypes and reproductive success (e.g., Price, 1970; Queller, 2017). For this purpose, it is immaterial that covariance is a purely correlational concept, blind to causality. To emphasize the crucial role of reproductive success in accounting for changes in gene frequencies, it seems natural to attach the label "fitness" to reproductive success. This then leads to the claim that a gene is positively selected if, and only if, it is positively *correlated* with fitness (provided that the correlation did not arise by chance). Let us summarize this view with the slogan: fitness is a *correlational predictor*. This view is expressed in the following quote (Birch, 2017b, p. 116):

> As Grafen (1982, 1984) emphasizes, it is a constraint on any fitness concept that if bearers of one allele are, on average, fitter than bearers of an alternative allele, then the former should be favoured by selection at the expense of the latter.

Note that the quote says nothing about causality. In particular, it does not account for the possibility that fitness averages may be subject to confounding factors.

But causality matters for the study of biological adaptation. Complex adaptations are sometimes called "engineering adaptations" (Lloyd, 2020) because they call for an engineering-style analysis of how they benefit their bearer. This creates a need to summarize an organism's relevant *causal properties* (i.e., those attributes which affect the outcomes of its interactions with its environment) with a suitable performance measure, traditionally called "fitness." Let us summarize this view with the slogan: fitness is a *performance measure*.

To aid our thinking about performance measures, we propose a metaphor. Envisage two types of machines, A and B, each designed to produce as many pencils as possible per year. This setup gives us prior knowledge of the relevant performance measure, which is the number of pencils produced. To establish which machine type is better fitted to the task, we must compare their performance under the same conditions;

otherwise, our statistical test will be inconclusive. For example, if machine type A, but not B, is regularly serviced by technicians, then the test will give a biased impression in favor of machine type A. Crucially, this bias is not a defect of the performance measure (i.e., the number of pencils), which we know to be correct. Instead, it is a symptom of the unsuitability of using averages to assess causal properties in the presence of confounders.

Let us now link this metaphor to evolution. Because IF_{folk} counts both provided and received benefits, a statistical association with IF_{folk} does not necessarily predict selection on a helping gene; i.e., IF_{folk} is not a correlational predictor. This is the double accounting problem of section 3B. Yet the machine metaphor shows that correlational data are the wrong benchmark to judge a performance measure's validity. In the metaphor, the technicians maintaining machine type A are analogous to relatives helping altruists in section 3B: In both cases, the received help confounds the performances of the entities we wish to compare. And in both cases, the proper response to the problem is not to blame the performance measure, but rather to use an appropriate method to assess causal properties. One such method is Pearl's (2009) do-calculus, which provides a mathematical analogue for the experimental practice of setting variables to predefined values by intervention. This method makes it possible to express an individual's causal effects as a function of its phenotype, allowing predictions based on causation rather than correlation (Figure 1 in Fromhage & Jennions, 2019).

What kind of fitness concept does Darwinism need?

Consider the classic example of the evolution of the vertebrate eye. Presumably, during the eye's evolution, there were many instances when an allele was selected for because it caused (on average, across the contexts in which it occurred) a favorable adjustment of some physical feature—say, the shape of the lens. A better-shaped lens would have rendered a sharper retinal image that facilitated the perception of features such as food, mates, predators, and relatives in need. Improved perception would, in turn, have caused possessors of the adjustment to have higher fitness than they would have had otherwise. Let us initially set aside the possibility of social interactions so that *LRS* is an appropriate fitness measure. The likely long-term outcome of many such adjustments, both to the eye and to other aspects of phenotypic design, is that successive generations of organisms acquired better eyesight, as well as other trait configurations whose causal effect increases the bearer's expected *LRS* compared to alternative configurations. (This trend toward *LRS*-enhancing traits should not be mistaken for a trend toward higher mean *LRS*—see Kokko (2021) and the section about population fitness below.) Note that in this fairly standard account of evolution, fitness is invoked as a *performance measure* that summarizes an organism's relevant causal properties. One could object to this account by saying that what really matters for an allele's propagation is its *correlation* with fitness, not its causal effect on it. But this would be beside the point because an allele that does not causally contribute to success gets us no closer to building a well-adapted organism. If we include social interactions, the above account remains essentially unchanged, except that the relevant performance measure is now IF_{folk}. One could then add to the previous objection (namely, that an allele's *correlation* with "fitness" is what really matters) that IF_{folk} cannot fill this role, even in principle, because IF_{folk} is not a *correlational predictor*. But this would still be beside the point for the same reason as before. If natural selection shapes (multi-) trait configurations based on the causal properties of the bearer, then characterizing these properties with a suitable fitness concept should be our priority. If this is incompatible with viewing fitness as a correlational predictor, then, in our view, a reasonable response is to stop viewing fitness as a correlational predictor.

There is no doubt about the immense usefulness of population and quantitative genetic theory, which invokes reproductive success (traditionally called "fitness") as a correlational predictor of genetic and phenotypic change. This impressive theory makes it tempting to think that, as with reproductive success in population and quantitative genetics, any valid fitness concept must be a correlational predictor. This inference, however, is unjustified. We have known at least since Hamilton (1964) that reproductive success is *not* the general design principle of adaptive evolution. Nor can we be certain, before identifying such a principle, what theoretical properties it may or may not have in common with reproductive success. We, therefore, cannot take for granted that the fitness concept embodying the design principle of adaptive evolution will be a correlational predictor. This realization pinpoints what was premature in Grafen's (1982) rejection of IF_{folk} based on the double accounting problem of section 3B: It was based on a tacit assumption that inclusive fitness must be a correlational predictor.

We speculate that the source of this confusion can be traced to the formerly uneasy relationship between mathematically oriented theory and the concept of causality. In the words of Pearl (2009):

> "It is an embarrassing yet inescapable fact that probability theory, the official mathematical language of many empirical sciences, does not permit us to express sentences such as "Mud does not cause rain"; all we can say is that the two events are mutually correlated, or dependent – meaning that if we find one, we can expect to encounter the other."

The same goes for statements such as: "the allele/trait causes (inclusive) fitness to increase."

Population fitness

Because individual-level adaptations may have positive, neutral, or negative consequences for conspecifics, natural selection is not expected to systematically enhance population-level properties such as population growth (Williams, 1966a). In technical terms, this means that evolution toward fitness-increasing traits neither implies nor requires the existence of a Lyapunov function (Devaney, 1986), i.e., a mathematical function that always increases along trajectories in the state space of the system. This point is sometimes missed in models where population growth (or another measure of "population fitness") is incidentally maximized as a result of "competitive" traits having been excluded a priori. This confusion has become enshrined in Sewell Wright's (1932) "adaptive landscape" metaphor, which characterizes natural selection as pushing toward "adaptive peaks" of high population mean fitness (for discussion, see Birch, 2015). It is worth pointing out that IF_{folk} is not susceptible to this source of confusion: because a given offspring may count toward the IF_{folk} of

multiple individuals, there is no biologically meaningful way to add up or average over the IF_{folk} values of all population members (Fromhage & Jennions, 2019, Q22). Thus, rather than being a defect of IF_{folk} (as suggested, e.g., by Dawkins, 1982, p. 185), this feature may serve as a safeguard against the misleading practice of focusing on population mean fitness.

Complex dynamics

Some authors have argued that the existence of complex dynamical phenomena (e.g., limit cycles, multiple and mixed equilibria) rules out the possibility of a general optimizing tendency in evolution (Allen & Nowak, 2016; Allen et al., 2013; Gintis, 2013). However, such phenomena merely rule out that evolution always favors the same phenotypes. The evaluation of individual adaptedness is relative to the environment in which it is being evaluated. Thus, when the present state of a population is itself an important component of the environment in which its members are selected, natural selection's optimizing tendency tracks a moving target that is continuously shifted by the evolutionary changes it induces (Fromhage & Houston, 2022; Maynard Smith, 1982; Odling-Smee et al., 2013). Although a population might thereby return to its initial state, indicating the absence of any enduring cumulative trend, this does not negate the importance of cumulative trends in nature. Even as some traits exhibit cyclical dynamics, others will likely face unidirectional selection. For example, in lizards with cyclical mating system dynamics due to frequency-dependent selection ("rock-paper-scissors": Zamudio & Sinervo, 2000), visual acuity and fast reflexes are probably consistently selected for. The ubiquity of organisms exhibiting "that perfection of structure and coadaptation which most justly excites our admiration" (Darwin, 1859, p. 3) suggests that cumulative improvement has occurred on a large scale.

Concluding remarks

When hearing the word "fitness," a population geneticist may think of a quantity used to predict gene-frequency change, a quantitative geneticist may think of a quantity used to predict phenotypic change, a behavioral ecologist may think of a performance measure for individuals, a theoretically oriented behavioral ecologist may think of an adaptedness criterion for strategies, and a philosopher, perhaps, may think of natural selection's improvement criterion. Although these ideas can complement each other (see section 10), depending on their precise interpretation they can also clash. In particular, if one commits to the view that fitness must be, above all else, a predictor of gene-frequency change, then this can derail the investigation of adaptive evolution in two ways. First, without careful reasoning about causality, it may motivate sweeping rejections of IF_{folk} that discourage exploring its full potential as a performance measure relevant to adaptive evolution (see the "double-accounting" problem of section 3B). Second, if we treat selection-driven gene-frequency change as synonymous with adaptive evolution, then we will miss the crucial distinction between adaptive and maladaptive cases of gene-level selection (Figure 1). To see the problem this creates, imagine taking a mixed bag of cases of adaptive and maladaptive evolution and then trying to find in this mixture a tendency for improvement that *always* holds. For example, because Hamilton's general rule (Queller, 1992; Gardner et al., 2011; also see Table 1) is designed to hold for positively selected genes, it also holds in Charlesworth's paradox for a positively selected nonhelping gene that decreases adaptedness at the phenotypic level. Thus, Hamilton's general rule is not a suitable diagnostic of adaptive evolution.

On the face of it, replacing $IF_{Hamilton}$ with IF_{folk} may appear to be, as Ågren (2021) put it, "quite a radical step." Yet we see this step as not only consistent with current biological practice but as providing a much-needed justification. We suspect that many Darwinians have long embraced IF_{folk} intuitively, either aware of its unorthodoxy or without wishing to draw attention to it. For example, in his book *Plan and Purpose in Nature*, George C Williams—a leading Darwinist of the 20th century—described inclusive fitness as "the overall ability of [an individual] to get her genes […] into future generations" (Williams, 1996, p. 60), without mentioning that anything is "stripped away." Without formally acknowledging such apparent departures from Hamilton's definition, however, the meaning of inclusive fitness is shrouded in ambiguity. Whether such ambiguity is acceptable in crucial aspects of evolutionary theory is a question we leave our readers to ponder.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

The simulation code is archived on Dryad: doi:10.5061/dryad.7sqv9s50r.

Author contributions

L.F. had the initial idea and wrote the first draft. M.D.J., L.M., and J.M.H. contributed by discussing ideas and writing.

Funding

J.M.H. was funded by the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 456626331. L.M. was funded by the Research Council of Finland (grant number 340130, awarded to Jussi Lehtonen).

Conflict of interest

Editorial processing of the manuscript was done independently of L.F. and J.M.H., who are Associate Editors of *Evolution*. The other authors declare no conflict of interest.

Acknowledgments

We thank Alasdair Houston, Jussi Lehtonen, Yagmur Erten, and David Haig for discussions and comments on the manuscript. We thank John Welch for his thoughtful review and for suggesting the summary in Box 1.

References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A. C., Andersson, M., Andre, J. B., Baalen, M. van, Balloux, F., Balshine, S., Barton, N., Beukeboom, L. W., Biernaskie, J. M., Bilde, T., Borgia, G., Breed, M., Brown, S., Bshary, R., Buckling, A., … Zink, A. (2011). Inclusive fitness theory and eusociality. *Nature*, *471*, E1–E4.

Ågren, J. A. (2021). *The gene's-eye view of evolution*. Oxford University Press.

Akcay, E., & Van Cleve, J. (2016). There is no fitness but fitness, and the lineage is its bearer. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *371*(1687), 20150085. https://doi.org/10.1098/rstb.2015.0085

Allen, B., & Nowak, M. A. (2016). There is no inclusive fitness at the level of the individual. *Current Opinion in Behavioral Sciences*, *12*, 122–128. https://doi.org/10.1016/j.cobeha.2016.10.002. Elsevier Ltd.

Allen, B., Nowak, M., & Wilson, E. O. (2013). Limitations of inclusive fitness. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(50), 20135–20139. https://doi.org/10.1073/pnas.1317588110

Ariew, A., & Lewontin, R. C. (2004). The confusions of fitness. *The British Journal for the Philosophy of Science*, *55*(2), 347–363. https://doi.org/10.1093/bjps/55.2.347

Bacon, F. (1623). *De dignitate et augmentis scientiarum*. Riegell and Wiessner.

Birch, J. (2015). Natural selection and the maximization of fitness. *Biological Reviews*, *91*(3), 712–727. https://doi.org/10.1111/brv.12190

Birch, J. (2016). Hamilton's two conceptions of social fitness. *Philosophy of Science*, *83*(5), 848–860. https://doi.org/10.1086/687869

Birch, J. (2017a). The inclusive fitness controversy: Finding a way forward. *Royal Society Open Science*, *4*(7), 170335. https://doi.org/10.1098/rsos.170335

Birch, J. (2017b). *The philosophy of social evolution*. Oxford University Press.

Birch, J. (2019). Inclusive fitness as a criterion for improvement. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *76*, 101186. https://doi.org/10.1016/j.shpsc.2019.101186

Birch, J., & Okasha, S. (2015). Kin selection and its critics. *Bioscience*, *65*(1), 22–32. https://doi.org/10.1093/biosci/biu196

Brandon, R. N. (1978). Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A*, *9*(3), 181–206. https://doi.org/10.1016/0039-3681(78)90005-5

Brandon, R. N. (2019). Natural selection. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Philosophy Department, Stanford University.

Brandon, R. N., & Beatty, J. (1984). The propensity interpretation of 'Fitness'—No interpretation is no substitute. *Philosophy of Science*, *51*, 342–347.

Byerly, H. C., & Michod, R. E. (1991). Fitness and evolutionary explanation. *Biology and Philosophy*, *6*(1), 1–22. https://doi.org/10.1007/bf02426816

Darwin, C. (1859). *On the origin of species by means of natural selection*. John Murray.

Darwin, C. (1860). *On the origin of species by means of natural selection* (2nd ed.). John Murray.

Davies, N. B. (2011). Cuckoo adaptations: Trickery and tuning. *Journal of Zoology*, *284*(1), 1–14. https://doi.org/10.1111/j.1469-7998.2011.00810.x

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Dawkins, R. (1982). *The extended phenotype*. Oxford University Press.

Dawkins, R. (1996). *Climbing mount improbable*. W.W. Norton & Company.

Devaney, R. L. (1986). *Introduction to chaotic dynamical systems*. Westview Press.

Dobzhansky, T. G. (1962). *Mankind evolving; The evolution of the human species*. Yale University Press.

Doebeli, M., Ispolatov, Y., & Simon, B. (2017). Towards a mechanistic foundation of evolutionary theory. *Elife*, *6*, 1–17.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon.

Fromhage, L., & Houston, A. I. (2022). Biological adaptation in light of the Lewontin–Williams (a)symmetry. *Evolution*, *76*(7), 1619–1624. https://doi.org/10.1111/evo.14502

Fromhage, L., & Jennions, M. D. (2019). The strategic reference gene: An organismal theory of inclusive fitness. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1904), 20190459. https://doi.org/10.1098/rspb.2019.0459

Garcia-Costoya, G., & Fromhage, L. (2021). Realistic genetic architecture enables organismal adaptation as predicted under the folk definition of inclusive fitness. *Journal of Evolutionary Biology*, *34*(7), 1087–1094. https://doi.org/10.1111/jeb.13795

Gardner, A. (2009). Adaptation as organism design. *Biology Letters*, *5*(6), 861–864. https://doi.org/10.1098/rsbl.2009.0674

Gardner, A., & West, S. A. (2010). Greenbeards. *Evolution*, *64*(1), 25–38. https://doi.org/10.1111/j.1558-5646.2009.00842.x

Gardner, A., West, S. A., & Wild, G. (2011). The genetical theory of kin selection. *Journal of Evolutionary Biology*, *24*(5), 1020–1043. https://doi.org/10.1111/j.1420-9101.2011.02236.x

Gayon, J. (1998). *Darwinism's struggle for survival*. Cambridge University Press.

Geritz, S. A. H., Kisdi, E., Meszena, G., & Metz, J. A. J. (1998). Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, *12*(1), 35–57. https://doi.org/10.1023/a:1006554906681

Gintis, H. (2013). Inclusive fitness and the sociobiology of the genome. *Biology and Philosophy*, *29*(4), 477–515. https://doi.org/10.1007/s10539-013-9404-0

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, *298*(5873), 425–426. https://doi.org/10.1038/298425a0

Grafen, A. (1984). Natural selection, kin selection and group selection. In: J. R. Krebs, & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 62–84). Blackwell Scientific Publications.

Grafen, A. (1985). A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology*, *2*, 28–89.

Grafen, A. (1999). Formal Darwinism, the individual-as-maximizing-agent analogy and bet-hedging. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *266*, 799–803.

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*, *238*(3), 541–563. https://doi.org/10.1016/j.jtbi.2005.06.009

Grafen, A. (2015). Biological fitness and the fundamental theorem of natural selection. *American Naturalist*, *186*(1), 1–14. https://doi.org/10.1086/681585

Grafen, A. (2018). The left hand side of the Fundamental Theorem of Natural Selection. *Journal of Theoretical Biology*, *456*, 175–189. https://doi.org/10.1016/j.jtbi.2018.07.022

Grafen, A. (2019). Should we ask for more than consistency of Darwinism with Mendelism? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *78*, 101224. https://doi.org/10.1016/j.shpsc.2019.101224

Grafen, A. (2020). The Price equation and reproductive value. *Philosophical Transaction of the Royal Society of London Series B: Biological Sciences*, *375*(1797), 20190356. https://doi.org/10.1098/rstb.2019.0356

Haig, D. (2020). *From Darwin to Derrida: Selfish genes, social selves, and the meanings of life*. MIT Press.

Haldane, J. B. S. (1938). *Heredity and politics*. W.W. Norton & Co.

Haldane, J. B. S. (1955). Population genetics. *New Biology*, *18*, 34–51.

Hamilton, W. D. (1964). Genetical evolution of social behaviour I. *Journal of Theoretical Biology*, *7*(1), 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of Ecology and Systematics*, *3*(1), 193–232. https://doi.org/10.1146/annurev.es.03.110172.001205

Hammerstein, P. (1996). Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology*, *34*(5-6), 511–532. https://doi.org/10.1007/BF02409748

Houston, A. I., Fromhage, L., & McNamara, J. M. (2023). A general framework for modelling trade-offs in adaptive behaviour. *Biological Review*, *99*(1), 56–69. https://doi.org/10.1111/brv.13011

Iseda, T. (1996). Changes in the concept of "fitness" in evolutionary biology. *Jissentetsugaku-Kenkyu (Studies Pract. Philos. - Japanese)*, *19*, 67–104.

Kokko, H. (2021). The stagnation paradox: The ever-improving but (more or less) stationary population fitness. *Proceedings of the*

Royal Society of London, Series B: Biological Sciences, *288*(1963), 20212145. https://doi.org/10.1098/rspb.2021.2145
Krimbas, C. B. (2004). On fitness. *Biology and Philosophy*, *19*(2), 185–203. https://doi.org/10.1023/b:biph.0000024402.80835.a7
Lee, J. J., & Chow, C. C. (2013). The causal meaning of Fisher's average effect. *Genetics Research*, *95*(2-3), 89–109. https://doi.org/10.1017/S0016672313000074
Lehmann, L., & Rousset, F. (2020). When do individuals maximize their inclusive fitness? *American Naturalist*, *195*(4), 717–732. https://doi.org/10.1086/707561
Leigh, E. G. (1971). *Adaptation and diversity*. Freeman, Cooper & Company.
Levin, S. R., Caro, S. M., Griffin, A. S., & West, S. A. (2019). Honest signaling and the double counting of inclusive fitness. *Evolution Letters*, *3*(5), 428–433. https://doi.org/10.1002/evl3.138
Levin, S. R., & Grafen, A. (2021). Extending the range of additivity in using inclusive fitness. *Ecology and Evolution*, *11*(5), 1970–1983. https://doi.org/10.1002/ece3.6935
Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia*, *118*, 63–82.
Lloyd, E. A. (2020). Units and levels of selection. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. The Metaphysics Research Lab, Philosophy Department, Stanford University.
Marrow, P., Johnstone, R. A., & Hurst, L. D. (1996). Riding the evolutionary streetcar: Where population genetics and game theory meet. *Trends in Ecology and Evolution*, *11*(11), 445–446. https://doi.org/10.1016/0169-5347(96)30036-0
Martens, J. (2019). Inclusive fitness as a measure of biological utility. *Philosophy of Science*, *86*(1), 1–22. https://doi.org/10.1086/701036
Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.
Maynard Smith, J. (1998). *Evolutionary genetics*. Oxford University Press.
McElreath, R., & Boyd, R. (2007). *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press.
McGraw, J. B., & Caswell, H. (1996). Estimation of individual fitness from life-history data. *American Naturalist*, *147*(1), 47–64. https://doi.org/10.1086/285839
Mellor, J. (2005). The dynamics of chromatin remodeling at promoters. *Molecular Cell*, *19*(2), 147–157. https://doi.org/10.1016/j.molcel.2005.06.023
Mills, S. K., & Beatty, J. (1979). The propensity interpretation of fitness. *Philosophy of Science*, *46*, 263–286.
Mylius, S. D., & Diekmann, O. (1995). On evolutionarily stable life histories, optimization and the need to be specific about density dependence. *Oikos*, *74*(2), 218–224. https://doi.org/10.2307/3545651
Neander, K. (1991). Functions as selected effects: The conceptual analyst's defense. *Philosophy of Science*, *58*(2), 168–184. https://doi.org/10.1086/289610
Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, *466*(7310), 1057–1062. https://doi.org/10.1038/nature09205
Odling-Smee, F. J., Laland, K. N., & Feldman, M. (2013). *Niche construction: The neglected process in evolution*. Princeton University Press.
Okasha, S. (2018). *Agents and goals in evolution*. Oxford University Press.
Okasha, S., & Martens, J. (2016). Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games. *Journal of Evolutionary Biology*, *29*(3), 473–482. https://doi.org/10.1111/jeb.12808
Patten, M. M., Schenkel, M. A., & Ågren, J. A. (2023). Adaptation in the face of internal conflict: The paradox of the organism revisited. *Biological Reviews*, *98*(5), 1796–1811. https://doi.org/10.1111/brv.12983
Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge University Press.
Price, G. R. (1970). Selection and covariance. *Nature*, *227*(5257), 520–521. https://doi.org/10.1038/227520a0
Price, G. R. (1972). Fisher's "fundamental theorem" made clear. *Annals of Human Genetics*, *36*(2), 129–140. https://doi.org/10.1111/j.1469-1809.1972.tb00764.x
Queller, D. C. (1992). A general model for kin selection. *Evolution*, *46*(2), 376–380. https://doi.org/10.1111/j.1558-5646.1992.tb02045.x
Queller, D. C. (1996). The measurement and meaning of inclusive fitness. *Animal Behaviour*, *51*(1), 229–232. https://doi.org/10.1006/anbe.1996.0020
Queller, D. C. (2017). Fundamental theorems of evolution. *American Naturalist*, *189*(4), 345–353. https://doi.org/10.1086/690937
Queller, D. C. (2019). What life is for: A commentary on Fromhage and Jennions. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20191060.
Ridley, M., & Grafen, A. (1981). Are green beard genes outlaws? *Animal Behaviour*, *29*(3), 954–955. https://doi.org/10.1016/s0003-3472(81)80034-6
Scott, T. W., & Wild, G. (2023). How to make an inclusive-fitness model. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *290*(2008), 20231310. https://doi.org/10.1098/rspb.2023.1310
Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus*. Chicago University Press.
Taylor, P. D., & Frank, S. A. (1996). How to make a kin selection model. *Journal of Theoretical Biology*, *180*(1), 27–37. https://doi.org/10.1006/jtbi.1996.0075
Taylor, P. D., Wild, G., & Gardner, A. (2007). Direct fitness or inclusive fitness: How shall we model kin selection? *Journal of Evolutionary Biology*, *20*(1), 301–309. https://doi.org/10.1111/j.1420-9101.2006.01196.x
van Veelen, M., Allen, B., Hoffman, M., Simon, B., & Veller, C. (2017). Hamilton's rule. *Journal of Theoretical Biology*, *414*, 176–230. https://doi.org/10.1016/j.jtbi.2016.08.019
Welch, J. J. (2023). The creativity of natural selection and the creativity of organisms: Their roles in traditional evolutionary theory and some proposed extensions. In T. E. Dickens, & B. J. A. Dickens (Eds.), *Evolutionary biology: Contemporary and historical reflections upon core theory* (pp. 65–107). Springer.
West, S. A., & Gardner, A. (2013). Adaptation and inclusive fitness. *Current Biology*, *23*(13), R577–R584. https://doi.org/10.1016/j.cub.2013.05.031
Wild, G., & Traulsen, A. (2007). The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, *247*(2), 382–390. https://doi.org/10.1016/j.jtbi.2007.03.015
Williams, G. C. (1966a). *Adaptation and natural selection*. Princeton University Press.
Williams, G. C. (1966b). Natural selection, the costs of reproduction, and a refinement of Lack's principle. *American Naturalist*, *100*, 687–690.
Williams, G. C. (1992). Gaia, nature worship and biocentric fallacies. *Quarterly Review of Biology*, *67*(4), 479–486. https://doi.org/10.1086/417796
Williams, G. C. (1996). *Plan and purpose in nature*. Phoenix.
Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In D. F. Jones (Ed.), *Proceedings of the sixth international congress of genetics* (Vol. *1*, pp. 356–366). Brooklyn Botanic Garden.
Zamudio, K. R., & Sinervo, B. (2000). Polygyny, mate-guarding, and posthumous fertilization as alternative male mating strategies. *Proceedings of the National Academy of Sciences*, *97*(26), 14427–14432. https://doi.org/10.1073/pnas.011544998