

**JYX**



**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Garcia-Costoya, Guillermo; Fromhage, Lutz

**Title:** Realistic genetic architecture enables organismal adaptation as predicted under the folk definition of inclusive fitness

**Year:** 2021

**Version:** Published version

**Copyright:** © 2021 The Authors. Journal of Evolutionary Biology published by John Wiley & Sc



**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

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>

# Realistic genetic architecture enables organismal adaptation as predicted under the folk definition of inclusive fitness

Guillermo Garcia-Costoya<sup>1,2</sup>  | Lutz Fromhage<sup>1</sup> 

<sup>1</sup>Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup>Department of Biology, University of Nevada, Reno, NV, USA

## Correspondence

Lutz Fromhage, Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, Jyväskylä 40014, Finland.

Email: lutz.fromhage@jyu.fi

## Abstract

A fundamental task of evolutionary biology is to explain the pervasive impression of organismal design in nature, including traits benefiting kin. Inclusive fitness is considered by many to be a crucial piece in this puzzle, despite ongoing discussion about its scope and limitations. Here, we use individual-based simulations to study what quantity (if any) individual organisms become adapted to maximize when genetic architectures are more or less suitable for the presumed main driver of biological adaptation, namely cumulative multi-locus evolution. As an expository device, we focus on a hypothetical situation called Charlesworth's paradox, in which altruism is seemingly predicted to evolve, yet altruists immediately perish along with their altruistic genes. Our results support a recently proposed re-definition of inclusive fitness, which is concerned with the adaptive design of whole organisms as shaped by multi-locus evolution, rather than with selection for any focal gene. They also illustrate how our conceptual understanding of adaptation at the phenotypic level should inform our choice of genetic assumptions in abstract simplified models.

## KEYWORDS

evolution of co-operation, natural selection, simulation, theory

## 1 | INTRODUCTION

A central idea in evolutionary theory is that natural selection shapes organisms through a process of cumulative improvement (Darwin, 1859; Dawkins, 1986). Although this is the widely accepted explanation for complex and well-adapted organismal design in nature, it has proven difficult to establish a general criterion for what qualifies as an improvement in adaptive design. This is a serious issue, for much of the theory's predictive and explanatory power hangs on its characterization of what properties of organisms are selected for. Given that most adaptive phenotypic change is underpinned by gene-frequency change (Fisher, 1930), we might expect organisms to evolve properties that help them spread their genes. But what are those properties? Hamilton (1964) made the brilliant insight that

an individual organism can propagate its genes not only through its own reproduction (direct fitness) but also by aiding its relatives that share the same genes (indirect fitness). He proposed that evolution favours phenotypes that maximize each organism's inclusive fitness, defined as follows (p. 8): 'Inclusive fitness may be imagined 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 [...]'.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Journal of Evolutionary Biology* published by John Wiley & Sons Ltd on behalf of European Society for Evolutionary Biology

After much debate (Abbot et al., 2011; Birch & Okasha, 2015; Nowak & Allen, 2015; Nowak et al., 2010) about the usefulness and limitations of Hamilton's inclusive fitness concept (henceforth  $IF_{\text{Hamilton}}$ ), Fromhage and Jennions (2019) advocated an alternative concept which does not involve any 'stripping' of fitness components as described in Hamilton's quote above. Instead, inclusive fitness is the sum of an individual's own offspring (including any accrued due to the behaviour/phenotype of relatives), plus its effects on its relatives' number of offspring, weighted by relatedness. They called this the 'folk definition of inclusive fitness' ( $IF_{\text{folk}}$ ) because it has been used informally by biologists for decades (Alcock, 2005; Mayr, 1991; Wilson, 1975; Zimmer & Emlen, 2016). The main advantage of  $IF_{\text{folk}}$  is that, unlike  $IF_{\text{Hamilton}}$ , it does not rest on the restrictive assumption that causal effects are additive (i.e. that the effects of an individual's actions are independent of the phenotypes of others; for discussion of additive causality at the level of organisms rather than alleles, see Birch, 2016, 2019). Another key feature of  $IF_{\text{folk}}$  is that its proposed merit as a phenotypic maximand (i.e. a quantity which organisms shaped by natural selection should appear designed to maximize) is based on the postulate that multi-locus long-term evolution tends to follow the genome's 'majority interest' (Hammerstein, 1996; Leigh, 1971). This postulate does not require that all positively selected genes conform to Hamilton's rule  $rb - c > 0$  (where  $r$  is relatedness and  $-c$  and  $b$  are changes caused to the reproduction of 'self' and 'other'), because short-term change driven by particular genes can be overturned in the long run by the combined effects of many genes (the so-called 'parliament of genes'; Leigh, 1971).

Despite encouraging signs that  $IF_{\text{folk}}$  is provoking a rethink (e.g. Queller, 2019), we have gained the impression from discussions with colleagues that many still struggle to fully grasp its meaning and import. For example a recurring objection is that, rather than being built on the notion of an *organism's inclusive fitness* (as represented by the definition of  $IF_{\text{Hamilton}}$  above), modern social evolution theory (e.g. Frank, 1997; Rousset, 2015) is built on the notion of a focal *allele's inclusive fitness effect* (a measure of selection at the gene level), and that there is hence no need to amend any perceived limitations of  $IF_{\text{Hamilton}}$ . While we agree that  $IF_{\text{Hamilton}}$  is indeed little used by modellers (cf. Q31 in Fromhage & Jennions, 2019), there are at least two reasons why focussing exclusively on selection at the gene level misses something important.

Firstly, the notion of optimal or 'adaptive' phenotypes is widely used by biologists, for example to generate testable hypotheses and to explain specific traits in terms of general principles. This notion hinges on the existence of a criterion of good phenotypic design. In particular, many biologists assume that an optimal phenotype is that which maximizes an individual's inclusive fitness – in the sense that feasible alternative traits and behaviours (e.g. when induced experimentally) would cause inclusive fitness to be lower. This notion concerns whole organisms and their causal effects, not particular genes. Hence, there is a gap between measures of gene-level selection and the ways in which inclusive fitness is actually used by biologists.

Secondly, since the true genetic complexity of adaptive evolution will usually defy full mathematical description, mathematical

analyses must make simplifying assumptions which from case to case may be more or less appropriate for modelling this process. In particular, if we find a given allele to be positively selected, there is no guarantee that the associated phenotypic change will be aligned with the Darwinian tendency of cumulative improvement which presumably we should aim to model. Yet unless we understand at a conceptual level what drives cumulative improvement, and what property is being improved, we have no means of judging which genetic assumptions succeed in capturing that process. Hence, calculations of gene-level selection are no substitute for identifying what phenotypic design principle (if any) is inherent in natural selection.

In the present study, we compare genetic architectures, implemented with individual-based simulations, to look at what quantity (if any) individual organisms become adapted to maximize. These genetic architectures vary in how well they capture the presumed main mechanism of biological adaptation, namely cumulative multi-locus evolution. As an expository device, we consider a classic example designed to illustrate the limitations of inclusive fitness.

## 2 | CHARLESWORTH'S PARADOX

McElreath and Boyd (2008) describe the following hypothetical situation inspired by Charlesworth's (1978) classic population-genetic study: 'Consider a species of bird in which young have the option of staying behind and helping their parents care for the next season's young rather than going out and trying to found their own nests. In this particular species, a situation arises each generation allowing an individual to sacrifice its own life to save the lives of four of its younger full siblings. Thus  $b = 4$ ,  $c = 1$ , and  $r = 0.5$ . According to Hamilton's rule, this behaviour should evolve ( $4 \times 0.5 > 1$ ), yet it cannot. In any individual in which the mutation arises, the allele will be destroyed.' Following conventional wisdom, McElreath and Boyd (2008, p. 99) identify as the solution to this paradox its violation of the 'weak selection' assumption presumed to be implicit in the notion of inclusive fitness: namely that Hamilton's rule (and, by extension, inclusive fitness) only applies for genes with a negligible effect on fitness. By contrast, Fromhage and Jennions (2019) have argued that their version of inclusive fitness theory applies to multi-locus evolution involving genes of various effect sizes, thus requiring no general assumption of 'weak selection'.

To revisit Charlesworth's paradox in the light of  $IF_{\text{folk}}$  theory, let us envisage a large (infinite) genetically uniform population in which the propensity for altruism is  $p$ . That is, each individual exhibits altruism with independent probability  $p$ . For simplicity and to match Charlesworth's assumption, let brood size  $N_{\text{brood}}$  be sufficiently large to neglect stochastic effects; in particular, we ignore the possibility that all offspring in a given brood might sacrifice their lives simultaneously, such that no sibling is left to benefit from the help. We follow Charlesworth's (1978) 'model 1' in assuming that each altruist provides  $b$  fitness units (in total) to its nonaltruistic siblings, which does not necessarily mean (contra McElreath & Boyd's interpretation) that helping saves the recipient's life. The total benefit of

altruism per brood is then given by  $pN_{\text{brood}}b$ , and the expected total benefit received per nonaltruistic sibling is.

$$B = \frac{pN_{\text{brood}}b}{N_{\text{brood}}(1-p)} = \frac{pb}{(1-p)} \quad (1)$$

Hence, the direct fitness of a nonaltruist is  $\text{baseline} + B$ , where  $\text{baseline}$  is the baseline fitness that would occur in the absence of help. According to Fromhage and Jennions (2019),  $IF_{\text{folk}}$  includes all of a focal individual's (relatedness-weighted) causal effects on its direct and indirect reproduction, compared to the counterfactual where the focal individual does not exist (as would occur, for example if the focal individual had died at conception). Accordingly, a nonaltruist's  $IF_{\text{folk}}$  is given by

$$IF_{\text{folk}} [\text{no altruism}] = \text{baseline} + B - rB \quad (2)$$

where the term  $-rB$  represents a fitness loss imposed on its siblings through participating in kin competition for available benefits. That is, if the focal nonaltruist had not existed, hence had not claimed any help for itself, then this help (of value  $B$ ) would instead have benefited some other sibling(s) (of relatedness  $r$ ). This formulation reflects the assumption that the focal nonaltruist's existence does not affect the number of helpers, who would have just helped someone else if the focal individual had been unavailable. Under this assumption, any given benefit obtained by a focal nonaltruist necessarily corresponds to an equal benefit withheld from others. An altruist's inclusive fitness is given by

$$IF_{\text{folk}} [\text{altruism}] = rb, \quad (3)$$

reflecting the assumption that its only causal effect (compared to its nonexistence) is to generate benefit  $b$  for a sibling. Fromhage and Jennions (2019) predicted that long-term evolution favours phenotypes yielding higher  $IF_{\text{folk}}$  such that, at evolutionary equilibrium,  $IF_{\text{folk}}$  should be maximized in the sense that no individual could increase its  $IF_{\text{folk}}$  if it unilaterally switched to an alternative phenotype. In the present case, this means that any stable mixture of altruists and nonaltruists (i.e. a mixed ESS (Maynard Smith, 1982)) must occur at a frequency of altruism where  $IF_{\text{folk}} [\text{no altruism}] = IF_{\text{folk}} [\text{altruism}]$ . Equalizing Equations (2) and (3), solving for  $p$  and adopting the convention of measuring  $b$  in units of  $\text{baseline}$  (implemented by setting  $\text{baseline} = 1$  without loss of generality), we thus obtain the evolutionarily stable frequency of altruism:

$$p = \frac{rb - 1}{b - 1} \quad (4)$$

This is the prediction based on  $IF_{\text{folk}}$  theory which we will now test using simulations. While the specific behaviour modelled is hypothetical, it is intended to represent the broad and ubiquitous category of adaptive 'goal directed' behaviours. Such behaviours tend to be complex in that they involve coordinated sensomotory routines that only plausibly arise by the Darwinian principle of cumulative improvement, rather than by a single chance mutation. Hence, when

studying the evolutionary origin of such behaviour, only simulations allowing for gradual, cumulative evolution – which, however, does not exclude strongly selected genes – have a reasonable a priori claim to biological realism.

### 3 | SIMULATIONS

We consider genetic systems that vary in the number and type of both 'primary' loci encoding the propensity for altruism, and 'regulatory' loci that may suppress the phenotype encoded by primary loci (see Figure S1 for a schematic). The simulations proceed in  $t_{\text{max}}$  discrete nonoverlapping generations, each following the flow chart given in Figure S2.

#### 3.1 | Genes and phenotypes

For simplicity, we assume haploid genetics (again following Charlesworth, 1978). Each individual has  $L_{\text{primary}}$  primary loci potentially coding for altruism and  $L_{\text{regulatory}}$  regulatory loci potentially coding for suppressing altruism. Loci are characterized by their allelic values, which are either discrete (0 or 1) or quantitative (between 0 and 1). In a given simulation, all loci of the same type (i.e. primary or regulatory) have the same type of allelic values (i.e. discrete or quantitative), but primary versus regulatory loci may differ in number and allelic type. Phenotypes are independently determined for each individual according to its genotypic propensity for altruism, which specifies its probability of behaving altruistically. An individual's propensity for altruism is its mean allelic value among primary loci, multiplied by the probability (equal to 1 minus the mean allelic value among regulatory loci) that regulatory genes will not suppress the altruistic phenotype. So, high allelic values at primary loci lead to a high altruism propensity unless counteracted by high allelic values at regulatory loci. In the special case with no regulatory loci ( $L_{\text{regulatory}} = 0$ ), altruism propensity is determined by primary loci alone. In the first generation, the population is initialized with all allelic values set to 0 for all primary and discrete regulatory loci. To speed up convergence, however, quantitative regulatory loci are initialized with randomly sampled values from a uniform distribution between 0 and 1.

#### 3.2 | Reproduction

In each time step,  $N_{\text{matings}}$  mating pairs are assembled by sampling individuals randomly with replacement (i.e. allowing individuals to mate multiply) according to their individual-specific mating propensity  $P_{\text{mating}}$  (see below). For simplicity, we do not model distinct sexes (i.e. individuals are hermaphrodites). Each pair produces  $N_{\text{brood}}$  offspring which inherit alleles by unlinked Mendelian inheritance. Mutations then occur independently at each locus (of each offspring) with probability  $\mu$ . When a gene mutates, its allelic value switches

from 0 to 1 or vice versa (for discrete loci), or a new allelic value is sampled from a uniform distribution between 0 and 1 (for quantitative loci). Then, phenotypes are determined as described above.

### 3.3 | Social interactions

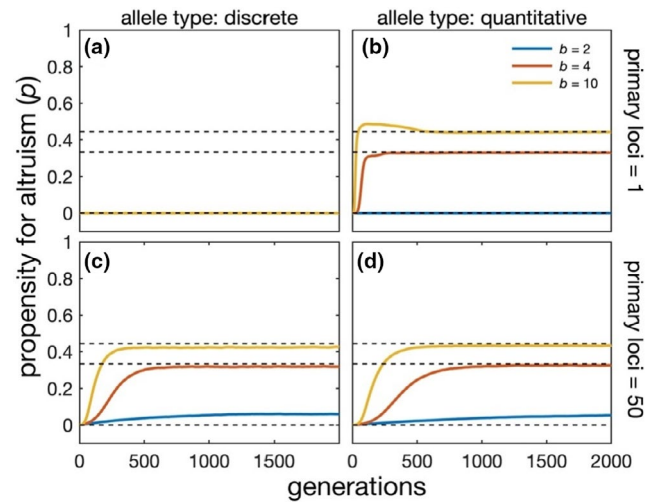
Because altruists are assumed to sacrifice their life to aid their siblings, they never enter the next generation. Instead, they each provide a benefit  $b$ , such that the total benefit per brood equals  $b$  times the number of altruists. This total benefit is then divided evenly among all nonaltruistic siblings within the same brood. When an individual receives a benefit of magnitude  $B$ , it is assigned mating propensity  $P_{\text{mating}} = 1 + B$ . Mating propensities are then used as sampling weights in determining mating success in the next generation (see above), which translates into expected reproductive success. For example, if an individual obtains a benefit of value  $B = 1$ , this corresponds to an increment of one unit of baseline fitness; so the individual's expected reproductive success is approximately doubled compared to what it would have obtained otherwise.

### 3.4 | Default settings

All simulations use  $N_{\text{matings}} = 10,000$  (but see Figure S3), thereby fixing population size, and  $\mu = 0.0001$  (but see Figure S4). The number of primary loci was set either to 1 or 50 (but see Figure S5), where the value 50 was chosen because it appears sufficient to model approximately continuous trait expression even when allelic values are discrete. The number of regulatory loci was set either to 0 or 1 or 50. We use  $N_{\text{brood}} = 10$  (but see Figure S6), which is sufficiently large to render unlikely the possibility of all offspring in a given brood expressing altruism at the same time. For example, with an altruism propensity of  $p = 0.44$  (the highest mean value observed in our simulations), the probability of all 10 offspring behaving altruistically was  $0.44^{10} = 0.0003$ .

## 4 | RESULTS

In most cases, altruism frequencies evolved to closely match predictions from  $IF_{\text{folk}}$  theory (Figures 1–3). The two notable exceptions occur when there is only one discrete primary locus and either no (Figure 1a), or one discrete regulatory locus (Figure 2a). Otherwise, our main results approximately match analytical predictions, and do so ever more closely as the mutation rate  $\mu$  is lowered, and as  $t_{\text{max}}$ ,  $N_{\text{matings}}$  and  $N_{\text{brood}}$  are increased (Figures S3–S6). In simulations differing only in initial allelic values at a quantitative regulatory locus, there is convergence to the predicted phenotype despite persistent differences at the genotypic level (Figure 3). Here, allelic values at the regulatory locus usually (but not always) stayed close to their initial level (Figure 3, first panel), while suitable complementary values evolved at the primary locus (Figure 3, second panel).



**FIGURE 1** Evolution of altruism under genetic systems without regulatory loci and (a) 1 discrete primary locus; (b) 1 quantitative primary locus; (c) 50 discrete primary loci; and (d) 50 quantitative primary loci. Coloured lines show mean values across 10 replicate simulations, for each of 3 levels of benefit  $b$ . Dashed lines indicate predictions from  $IF_{\text{folk}}$  theory (calculated from eq. 4 with  $r = 0.5$ ). Evolved altruism levels closely match predictions in (b)–(d) but not (a)

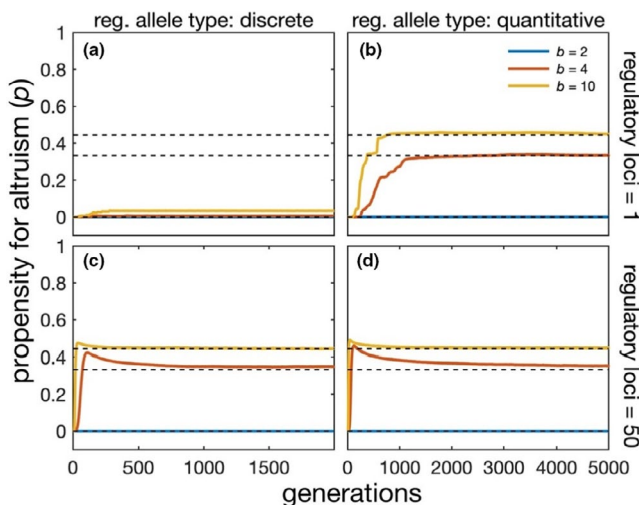
## 5 | DISCUSSION

We mostly found close agreement between our simulations and predictions of  $IF_{\text{folk}}$  theory (Figures 1–3). The two notable exceptions are cases with only one or two loci with discrete alleles (Figures 1a and 2a). In order to understand these results, envisage a situation where a rare altruism allele arises in an otherwise selfish population. The causal effect of expressing this allele (compared to the counterfactual of not expressing it) is to generate, at cost  $c = 1$  to the focal individual, an indirect benefit (say,  $b = 4$ ) in a sibling ( $r = 0.5$ ). This amounts to a positive net effect ( $0.5 \times 4 - 1 > 0$ ) on the focal individual's  $IF_{\text{folk}}$ , indicating that an increase in the propensity for altruism is in the genome's 'majority interest' and should be selected for in the long run under multi-locus evolution. This argument invokes what Fromhage and Jennions (2019) called 'Hamilton's phenotypic rule', which provides a criterion for whether a given phenotypic change increases a focal individual's  $IF_{\text{folk}}$ . Note that, whereas  $IF_{\text{folk}}$  evaluates the causal effects of an organism as a whole (compared to its absence), Hamilton's phenotypic rule evaluates the causal effect of a given phenotypic change (compared to no change). So, if phenotypic changes that satisfy Hamilton's phenotypic rule tend to be selected for, then whole organisms should tend to become adapted to maximize their  $IF_{\text{folk}}$ .

In contrast, under the assumptions of Figure 1a (with one discrete primary locus), selection favours the nonaltruistic allele. This occurs because the altruistic allele is always expressed (i.e. has full penetrance), which, combined with the fact that only nonaltruists receive help, biases the flow of social benefits towards nonaltruistic genotypes. The nonaltruistic allele therefore qualifies as what Fromhage and Jennions' called a 'mirror effect rogue gene'. This is an allele that

reduces the  $IF_{\text{folk}}$  of the organisms expressing it, but is still selected for due to the biased flow of social benefits arising under high levels of penetrance. However, consistent with the prediction that *mirror effect rogue genes* are of little importance in multi-locus evolution (Fromhage & Jennions, 2019), altruism readily evolves when the rigid genetic constraints operating in Figures 1a and 2a are relaxed. This is made possible by genes having imperfect penetrance, which can realign the flow of social benefits with the genome's 'majority interest'. Further rogue genes can then no longer invade, indicating that they are best viewed as a manifestation of a genetic constraint rather than of a genetic conflict (Queller, 2019).

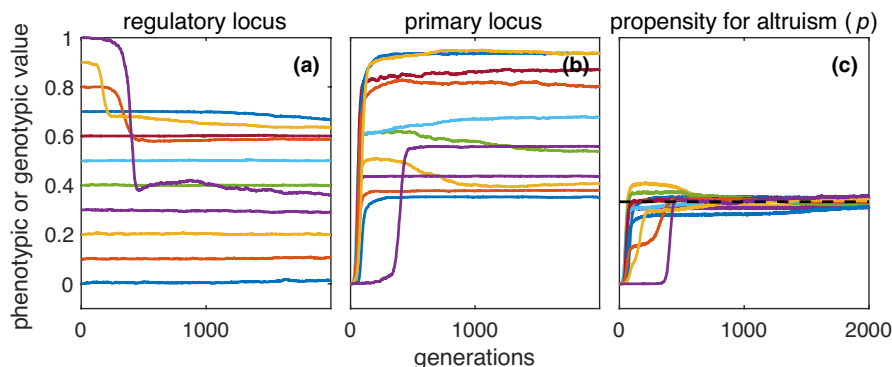
We also found that alternative regulatory mechanisms can evolve to produce the same optimal phenotype (Figure 3), such that high levels of altruism propensity (as encoded by primary loci)



**FIGURE 2** Evolution of altruism under genetic systems with one discrete primary locus and (a) 1 discrete regulatory locus; (b) 1 quantitative regulatory locus; (c) 50 discrete regulatory loci; and (d) 50 quantitative regulatory loci. Coloured lines show mean values across 10 replicate simulations, for each of 3 levels of benefit  $b$ . Dashed lines indicate predictions from  $IF_{\text{folk}}$  theory (calculated from eq. 4 with  $r = 0.5$ ). Evolved altruism levels closely match predictions in (b)–(d) but not (a)

select for high levels of altruism suppression and vice versa. This illustrates the classic idea, sometimes called the 'genetic theory of relativity' (Mayr, 1963), that a gene's selective value depends on its genetic background. This idea has sometimes been used (e.g. Gould, 2002; Sober, 1984) to challenge the 'gene selectionist' view that gene-level selection holds the key to understanding phenotypic evolution (Williams, 1966; Dawkins, 1976; for a recent defence, see Queller, 2020). In the light of  $IF_{\text{folk}}$  theory in general and Figure 3 in particular, we argue that – notwithstanding its gene-level underpinnings – the 'big picture' of what is selected for at the organism level is not fully captured by gene-level considerations. Focussing on organisms pinpoints the design principle that guides phenotypic evolution (regardless of many genetic details: e.g. Figure 3), and this allows us to draw a crucial distinction between cases where this principle can or cannot operate (e.g. Figure 1a). Once this distinction is made, it becomes easy to see that selection for genes to increase their number of copies is not the same thing as selection for organisms to maximize their inclusive fitness (contra, e.g. Dawkins, 1978). Of course, at some level, this is common knowledge. For example, although almost any maladaptive trait could evolve if encoded by a segregation-distorter gene, nobody takes that to imply that we cannot predict anything about phenotypic evolution. Instead, there is (implicit) agreement that that's just the wrong kind of genetic assumption to make for studying adaptive evolution. But if that is the case, then what is the right kind of genetic assumption? In particular, is assuming fair meiosis sufficient to ensure biologically plausible outcomes? We think not, but invite our readers to make up their own minds in the light of our present findings.

To eliminate a source of ambiguity, we note that there exist different types of evolutionary explanations that are relevant in this context. Rosales (2005), drawing on Sober (1983), distinguishes historical explanations, which show how an event to be explained was produced through a specific sequence of earlier events, from 'modal' explanations, which show how an event would have occurred regardless of which actual history transpired among a range of possibilities. We do not doubt that gene-level selection (in combination with chance and mutation pressure) can provide historical explanations



**FIGURE 3** Evolution of altruism with 1 quantitative primary locus and 1 quantitative regulatory locus, the latter being set to different starting values (evenly spaced between 0 and 1) for each replicate simulation (coloured lines). In the rightmost panel, the dashed line at  $p = 0.33$  indicates the prediction from  $IF_{\text{folk}}$  theory (calculated from eq. 4 with  $r = 0.5$  and  $b = 4$ ). Evolved altruism levels closely match the prediction despite large enduring differences at the genotype level

for all our simulation outcomes. Yet to also explain why genetic details don't matter for (say) the converging outcomes of Figure 3c, we need to invoke improvement of whole-organism design as their unifying feature. Likewise, when seeking to explain real-world adaptations, even full knowledge of genetic history (which in any case is forbiddingly hard to come by) can be no substitute for identifying the general forces and principles which have shaped that history. In our view, recognizing the complementarity of these explanatory approaches makes the difference between 'good' reductionism, which means explaining a thing in terms of what it reduces to (e.g. its parts and their interactions), and what Dennett (1995) called 'greedy' reductionism, which occurs when trying to skip whole layers or levels of theory. We should add that these remarks take nothing away from the value of using the 'gene's eye view' (Ågren, 2021; Dawkins, 1976) as a thinking tool for generating evolutionary hypotheses.

Those long convinced that adaptation leads to increased inclusive fitness will rightly ask what is gained by replacing the traditional  $IF_{\text{Hamilton}}$  concept with  $IF_{\text{folk}}$ . To answer this question, compare the predicted altruism frequency of Equation (4) with a similar prediction based on  $IF_{\text{Hamilton}}$ . We obtain the latter prediction by proceeding as before, except that we now exclude from eq. 2 the term (+ B) representing received benefits that 'can be considered as due to the individual's social environment' (Hamilton, 1964, p. 8). (For other ways to interpret Hamilton's 'stripping procedure', which would not change our conclusions, see Fromhage & Jennions, 2019, section 7). This yields  $p = 1 - rb$  as the predicted frequency of altruism, amounting to the biologically absurd prediction that altruism should decrease when it becomes more beneficial. This prediction arises because a nonaltruist's negative causal effect on its siblings (through participating in kin-competition for available benefits) increases with  $b$ , while the counteracting advantage of receiving more benefits is excluded. One could object to this calculation by pointing out that  $IF_{\text{Hamilton}}$  is committed to an additive model of causality (Birch, 2016, 2019), which contradicts our assumption that a focal individual's phenotype can affect what it receives from others. We agree, but still wish to illustrate how badly  $IF_{\text{Hamilton}}$  can fail when its restrictive assumptions do not hold, as is often the case in nature.

We emphasize that our prediction based on  $IF_{\text{folk}}$  fully agrees with several standard methods of evolutionary modelling. Namely, our eq. 4 corresponds to Charlesworth's (1978) eq. 25 which was derived from a (weak selection) population-genetic model (where in Charlesworth's notation, our parameters  $p$  and  $b$  correspond to  $p^*/(p^* + 1)$  and  $k$ , respectively, and  $r = 0.5$  is implicit), and it can also be recovered with Taylor and Frank's (1996) 'direct fitness method'. This may seem surprising considering that especially the latter method is often discussed in a context of  $IF_{\text{Hamilton}}$  maximization. However, since these methods focus on selection for incremental phenotypic change – without directly attending to Hamilton's 'stripping procedure' – there is in fact nothing contradictory about the finding that they predict an equilibrium at which organisms turn out to maximize  $IF_{\text{folk}}$  rather than  $IF_{\text{Hamilton}}$ .

Inspired by common simplifying assumptions of theoretical models, many discussions of inclusive fitness theory invoke a false

dichotomy between 'weak' and 'strong' selection, while neglecting the much more realistic possibility that genes of various effect sizes (hence subject to various strengths of selection) jointly affect the course of evolution. This has led to a widespread view that selection must generally be weak for inclusive fitness arguments to apply. Fromhage and Jennions (2019) reject this claim, and our results support their view: genes of various effect sizes (including large effects) acting together select for  $IF_{\text{folk}}$  – maximizing behaviour (Figures 1b, 2b-d, and 3).

Recently, Lehmann and Rousset (2020) conducted a population-genetic analysis about when individuals should maximize their inclusive fitness. They concluded that the essential role of weak selection 'precludes [...] a general rational actor-centered representation of adaptation'. Yet that conclusion must be viewed in light of the fact that their study neglected the possibility of simultaneous occurrence of genes with various effect sizes, as opposed to what we present here. Moreover, Lehmann and Rousset's version of (organismal) inclusive fitness (their Equation 4) did not actually measure the combined direct and indirect reproductive success of an organism as a whole. What is normally meant by the inclusive fitness of an individual organism captures 'the effects of his lifetime's set of deeds [as compared to] a hypothetical lifetime of total inaction – as though he had never been conceived'; Dawkins, 1982, p. 186). By contrast, while Lehmann and Rousset's direct fitness component implicitly includes all of an individual's own offspring (in agreement with  $IF_{\text{folk}}$ ), their indirect fitness component includes only the marginal effect 'stemming from a single [...] gene copy switching to expressing a copy of the [weakly selected] 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 kin-selected adaptations) are excluded! This makes Lehmann and Rousset's maximand neither an absolute nor a marginal measure of reproductive success (cf. Q18 in Fromhage & Jennions, 2019), but rather a mixture of both: it is absolute with regard to direct fitness, yet marginal with regard to indirect fitness. In our view, despite being mathematically valid (as are countless other functions that are all maximized by the same phenotype; cf. Q26 in Fromhage & Jennions, 2019), this 'half-marginal' formulation of  $IF$  seems needlessly counter-intuitive. For example, it implies that a sterile worker ant's  $IF$  would be approximately zero – give or take some weak effect of a mutant allele; and that a worker bee's  $IF$  would approximate her direct fitness from laying the occasional unfertilized egg, despite her phenotypic design being overwhelmingly dedicated to obtaining indirect fitness. It is worth adding that, although Hamilton's (1964) mathematical model made no explicit distinction between an allele's marginal effect and an organism's overall effect (due to his simplifying assumption that all of an organism's social effects were attributable to a single allele), Hamilton's verbal definition suggests that he invented the concept of inclusive fitness specifically for the purpose of capturing the latter aspect. If this is so, then we think  $IF_{\text{folk}}$  preserves the spirit of Hamilton's idea better than Lehmann and Rousset's half-marginal formulation does.

A similar half-marginal formulation of *IF*, with similar drawbacks, was proposed by Levin and Grafen (2021). Their version sums three components, namely '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  $y$  instead of the 'incumbent strategy'  $x$  played by almost all other individuals in the population. Like Lehmann & Rousset's version, this 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 effects. In addition, it faces the difficulty that 'baseline asocial fitness' is merely a theoretical abstraction that cannot be measured in practice. (To see this, ask yourself what a concrete biological example might look like of an organism having no effects whatsoever on its relatives. Should it disperse so far as to never meet, nor indirectly compete with, any relatives? No, because this very act of dispersal would itself qualify as a social act, in the sense of making a difference to the relatives in question. Hence, in practice, there may be no way for an organism to completely avoid having social effects of one kind or another.) Apart from its 'asocial' component – which in any case is inessential while assumed constant – Levin & Grafen's *IF* boils down to a criterion for the pairwise comparison of phenotypes. As such, it cannot be interpreted as a measure of individual performance which 'each organism appears to be attempting to maximise' (Hamilton, 1964, p. 1) by means of its causal effects on the world.

In conclusion, if  $IF_{\text{folk}}$  correctly identifies the organismal design principle guiding cumulative improvement in nature, then our finding that it cannot operate under overly simplistic assumptions is a cautionary note for modellers. We can hardly expect models to provide valid biological insights about long-term evolution unless their assumptions are informed by our best conceptual understanding of how adaptative evolution works (Queller, 2019). At the same time, the finding that gene-level selection sometimes opposes natural selection's improving tendency stymies any theoretical ambitions to prove, in the style of (some interpretations of) Fisher's fundamental theorem (Fisher, 1930; Grafen, 2015), that natural selection will unfailingly tend to improve organismal design whenever there is additive genetic variance.

## ACKNOWLEDGMENTS

We thank Jussi Lehtonen for alerting us to Charlesworth's paradox and for helpful comments on the manuscript; and Michael Jennions, Alan Grafen and David Queller for helpful comments on the manuscript.

## CONFLICT OF INTEREST

We declare no competing interests.

## AUTHOR CONTRIBUTIONS

GGC wrote the first draft of both the manuscript and the simulation code. LF had the idea and contributed to writing and coding.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/jeb.13795>.

## OPEN RESEARCH BADGES



This article has been awarded <Open Materials, Open Data> Badges. All materials and data are publicly accessible via the Open Science Framework at [<https://datadryad.org/stash/share/6wliYDR-HwV-i7cCkFagb05TcoRfs2KhyJW-JFobKqY>].

## DATA AVAILABILITY STATEMENT

The used MATLAB code can be found in Dryad at <https://datadryad.org/stash/share/6wliYDR-HwV-i7cCkFagb05TcoRfs2KhyJW-JFobKqY>.

## ORCID

Guillermo Garcia-Costoya  <https://orcid.org/0000-0003-0522-6108>

Lutz Fromhage  <https://orcid.org/0000-0001-5560-6673>

## REFERENCES

- Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A. C., Andersson, M., Andre, J. B., van Baalen, M., 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 Univ. Press.
- Alcock, J. (2005). *Animal behavior: An evolutionary approach*. Sinauer Associates.
- Birch, J. (2016). Hamilton's two conceptions of social fitness. *Philosophy of Science*, 83, 848–860. <https://doi.org/10.1086/687869>
- Birch, J. (2019). Inclusive fitness as a criterion for improvement. *Studies in History and Philosophy of Science Part C: 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, 22–32. <https://doi.org/10.1093/biosci/biu196>
- Charlesworth, B. (1978). Some models of the evolution of altruistic behaviour between siblings. *Journal of Theoretical Biology*, 72, 297–319. [https://doi.org/10.1016/0022-5193\(78\)90095-4](https://doi.org/10.1016/0022-5193(78)90095-4)
- Darwin, C. (1859). *On the origin of species by means of natural selection*. John Murray.
- Dawkins, R. (1976). *The selfish gene*. Oxford University Press.
- Dawkins, R. (1978). Replicator selection and the extended phenotype. *Zeitschrift für Tierpsychologie*, 47, 61–76.
- Dawkins, R. (1982). *The extended phenotype*. Oxford University Press.
- Dawkins, R. (1986). *The blind watchmaker*. Norton.
- Dennett, D. (1995). *Darwin's dangerous idea*. Simon & Schuster.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Clarendon.
- Frank, S. A. (1997). The price equation, Fisher's fundamental theorem, Kin selection, and causal analysis. *Evolution*, 51, 1712.
- 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>
- Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard University Press.



- Grafen, A. (2015). Biological fitness and the fundamental theorem of natural selection. *The American Naturalist*, 186, 1–14. <https://doi.org/10.1086/681585>
- Hamilton, W. D. (1964). Genetical evolution of social behaviour I. *Journal of Theoretical Biology*, 7, 1–16.
- Hammerstein, P. (1996). Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology*, 34, 511–532.
- Lehmann, L., & Rousset, F. (2020). When do individuals maximize their inclusive fitness? *The 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., & Grafen, A. (2021). Extending the range of additivity in using inclusive fitness. *Ecology and Evolution*, 11(4), 1970–1983.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.
- Mayr, E. (1963). *Animal species and evolution*. Harvard University Press.
- Mayr, E. (1991). *One long argument. Charles Darwin and the genesis of modern evolutionary thought*. Harvard University Press.
- McElreath, R., & Boyd, R. (2008). *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press.
- Nowak, M. A., & Allen, B. (2015). Inclusive fitness theorizing invokes phenomena that are not relevant for the evolution of eusociality. *PLoS Biology*, 13, e1002134. <https://doi.org/10.1371/journal.pbio.1002134>
- Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466, 1057–1062. <https://doi.org/10.1038/nature09205>
- Queller, D. C. (2019). What life is for: A commentary on Fromhage and Jennions. *Proceedings of the Royal Society B-Biological Sciences*, 286, 20191060. <https://doi.org/10.1098/rspb.2019.1060>
- Queller, D. C. (2020). The gene's eye view, the Gouldian knot, Fisherian swords and the causes of selection. *Philosophical Transactions of the Royal Society of London. Series B*, 375, 20190354.
- Rosales, A. (2005). John Maynard Smith and the natural philosophy of adaptation. *Biology and Philosophy*, 20, 1027–1040. <https://doi.org/10.1007/s10539-005-9021-7>
- Rousset, F. (2015). Regression, least squares, and the general version of inclusive fitness. *Evolution*, 69, 2963–2970.
- Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43, 201–210. <https://doi.org/10.1007/BF00372383>
- 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, 27–37. <https://doi.org/10.1006/jtbi.1996.0075>
- Williams, G. C. (1966). *Adaptation and natural selection*. Princeton University Press.
- Wilson, E. O. (1975). *Sociobiology*. Harvard University Press.
- Zimmer, C., & Emlen, D. (2016). *Evolution - making sense of life* (2nd ed.). W.H. Freeman and Co.

### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Garcia-Costoya G, Fromhage L. Realistic genetic architecture enables organismal adaptation as predicted under the folk definition of inclusive fitness. *J Evol Biol*. 2021;34:1087–1094. <https://doi.org/10.1111/jeb.13795>