

# There is no inclusive fitness at the level of the individual

## Benjamin Allen<sup>1,2</sup> and Martin A Nowak<sup>2,3,4</sup>

Inclusive fitness theory attempts to explain the evolution of social behavior at the level of the individual. The inclusive fitness of an individual is defined as the sum of all the effects this individual has on itself and others, with each term multiplied by a relatedness coefficient. However, this quantity does not exist for most evolutionary processes, because biological interactions are typically nonlinear and fitness effects are not additive. Six recent publications, by proponents of inclusive fitness theory and outside observers, agree with our criticism on two mathematical points: in general, there is no inclusive fitness at the level of the individual and natural selection does not lead organisms to maximize inclusive fitness. Here we review the current state of the controversy over inclusive fitness, highlight points of mathematical agreement, and discuss prospects for a way forward.

### Addresses

<sup>1</sup> Department of Mathematics, Emmanuel College, Boston, MA 02115, United States

<sup>2</sup> Program for Evolutionary Dynamics, Harvard University, Cambridge, MA 02138, United States

<sup>3</sup> Department of Mathematics, Harvard University, Cambridge, MA 02138, United States

<sup>4</sup> Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, United States

Corresponding author: Allen, Benjamin ([benjcallen@gmail.com](mailto:benjcallen@gmail.com))

Current Opinion in Behavioral Sciences 2016, 12:122–128

This review comes from a themed issue on Behavioral ecology

Edited by Andrew Sih and Alex Kacelnik

<http://dx.doi.org/10.1016/j.cobeha.2016.10.002>

2352-1546/© 2016 Published by Elsevier Ltd.

## Introduction

The evolution of cooperation has fascinated biologists since Darwin. JBS Haldane, when asked if he would sacrifice his life by jumping into a river to save a drowning brother, famously remarked ‘No, but I would to save two brothers, or eight cousins’. The implicit idea is that a gene — say, for saving a relative — can spread indirectly, by aiding copies of itself in other individuals. In this way, cooperative behaviors can evolve even if they are costly to the individual performing them.

How can this insight be formalized mathematically? The straightforward way is to suppose there is a

mutation that induces the behavior Haldane is contemplating — or at least increases the likelihood of that behavior in some environments. Obviously there need not be a single gene for saving your brother, but there must be a genetic foundation for the trait. The next step is to formulate an evolutionary model in order to understand the circumstances under which natural selection favors this mutation (Figure 1). This approach gives us conditions for the success of this mutation, telling us what to measure empirically and how to interpret these measurements. The approach can also be extended to study cases where multiple mutations are required for a particular trait or for the phenotypic plasticity toward a trait.

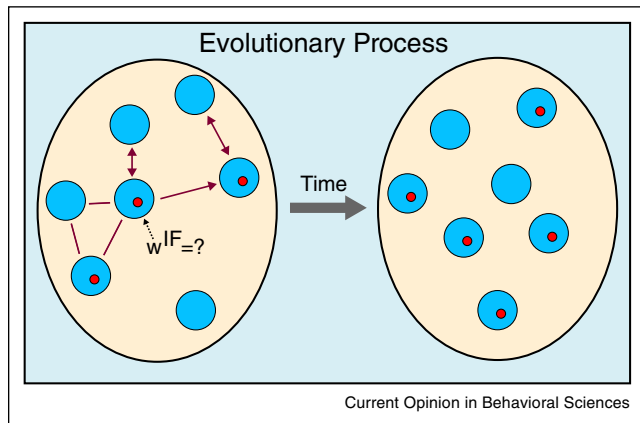
This framework of analysis constitutes a gene-level, evolutionary dynamics approach. It is mathematically and conceptually complete, and captures both direct and indirect genetic reproduction. But it is also appealing to instead try to understand evolution in terms of individual-level adaptations. After all, the original theory of evolution, before its genetic underpinning was discovered, focused on understanding how individual organisms become adapted to their environment.

Individual adaptedness can be quantified by *fitness*, defined as expected number of adult offspring. Over successive generations, natural selection could favor genes whose bearers have the highest fitness. To illustrate this principle, consider an asexual population of fixed size with two genetic types labeled 0 and 1. Denote by  $g_i$  and  $w_i$  the genetic type and fitness, respectively, of individual  $i$  in a given generation. From the current generation to the next, the abundance of type 1 changes from  $\sum_i g_i$  to  $\sum_i w_i g_i$ . Thus type 1 increases if its average fitness,  $\sum_i w_i g_i / \sum_i g_i$ , exceeds 1.

If an individual’s fitness depends only on its own genotype, selection favors the genotype with the highest fitness. This insight encapsulates the principle of survival of the fittest. It moreover suggests that evolution should lead organisms to act as if maximizing their fitness. However, fitness maximization does not apply to social behavior, whereby organisms affect each other’s fitness. Haldane does not maximize his fitness by sacrificing himself to save his brothers.

Hamilton [1] attempted to extend the fitness-maximization principle to social behavior, by replacing fitness with *inclusive fitness*. To define inclusive fitness, he made a restrictive assumption: each individual contributes a particular additive amount to the fitness of each other. In

Figure 1



A mathematical model of evolutionary dynamics specifies how a population changes over time. A mutant gene could affect the fitness of its bearer, as well as the fitnesses of others some of who might carry copies of the gene. Interactions can be multilateral and non-additive. Mathematical analyses of evolutionary dynamics give conditions under which a mutation affecting social behavior is favored by natural selection. This approach provides a complete description of evolutionary dynamics. In contrast, inclusive fitness theory attempts to define the quantity ‘inclusive fitness’ of an individual. Mathematical analysis has shown that this quantity does not exist in general but only for special cases. Moreover, there is no need to calculate such a quantity: every phenomenon in evolutionary biology can be fully analyzed without ever invoking or calculating ‘inclusive fitness’.

our hypothetical example, this means that each individual  $i$  with genotype  $g_i = 1$  changes the fitness of each individual  $j$  (including itself) by an amount  $e_{ij}$ . The fitness  $w_i$  of each individual  $i$  can then be written as

$$w_i = w_0 + \sum_{j=1}^N e_{ij}g_j. \tag{1}$$

Here  $w_0$  represents baseline fitness. Fitness according to Eqn 1, which is sometimes called *direct* or *neighbor-modulated* fitness, is recipient-centered: the consequences of social behaviors are grouped according to their recipient. Hamilton’s idea was to re-assign fitness effects to the actor, resulting in the quantity of inclusive fitness:

$$w_i^{IF} = w_0 + \sum_{j=1}^N e_{ij}g_j. \tag{2}$$

Although  $w_i$  and  $w_i^{IF}$  may differ for an individual, they are the same at the population level in the sense that  $\sum_i w_i g_i = \sum_i w_i^{IF} g_i$ . Thus, the frequency of type 1 increases, if its average inclusive fitness  $\sum_i w_i^{IF} g_i / \sum_i g_i$  exceeds 1.

Under additional assumptions (no dominance, fair meiosis), the above argument extends to sexually reproducing populations. With further assumptions still — that selection is weak and that the population structure is

fixed in some sense [2] — the genetic variables  $g_i$  can be replaced by suitably-defined relatedness coefficients  $r_{ij}$ , yielding

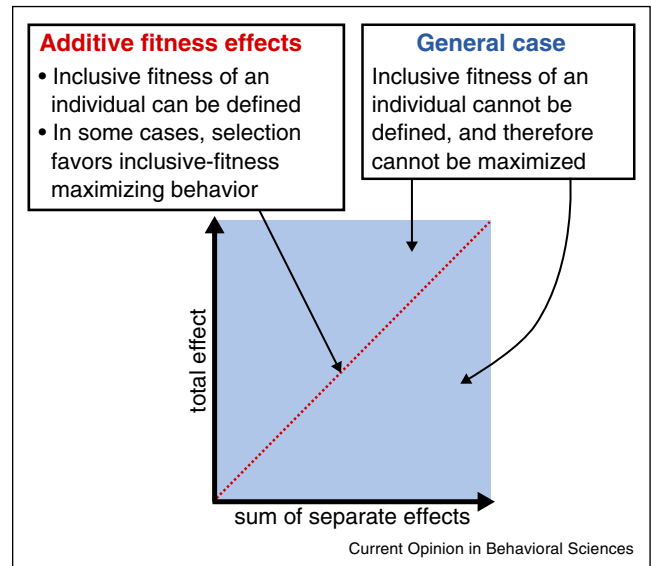
$$w_i^{IF} = w_0 + \sum_{j=1}^N e_{ij}r_{ij} \tag{3}$$

While Eqn 2 applies in a specific state, Eqn 3 applies to the overall selection process, provided that the additional assumptions hold.

Inclusive fitness has conceptual appeal as a target of selection that could be under an individual organism’s control. This feature has been interpreted as a design principle: ‘Natural selection explains the appearance of design in the living world, and inclusive fitness theory explains what this design is for. Specifically, natural selection leads organisms to become adapted as if to maximize their inclusive fitness’ [3].

But do the mathematics justify such sweeping interpretations? After all, to define the inclusive fitness of an individual, one must assume that all individuals contribute separate, well-defined amounts of fitness to each other.

Figure 2



In order to define the inclusive fitness of an individual, one must assume that individuals contribute well-defined amounts of fitness to each other. Under this assumption, each individual’s fitness is equal to a sum of separate contributions due to each individual (including itself), with no synergistic or nonlinear interactions. If this is the case, then inclusive fitness can be defined as in Eqn 2. Under additional assumptions [7\*\*], selection favors individual behaviors that maximize inclusive fitness. For general (or generic) evolutionary processes, however, there is no inclusive fitness of an individual [2,8\*\*,9], and individuals either do not maximize any quantity [7\*\*], or maximize a quantity that is not inclusive fitness [10,7\*\*,11\*\*,12].

It is easy to envisage biological interactions that violate this assumption. For example, it might take two individuals working together to save Haldane's brother from drowning. Or the reproductive rate of an insect colony could be a nonlinear function of the number of workers. Experiments have shown that nonlinear and non-additive interactions are ubiquitous [4–6]. Therefore, for a general biological scenario, the inclusive fitness of an individual cannot be defined, does not exist and cannot be maximized (Figure 2).

This limitation of inclusive fitness theory has been pointed out numerous times [13–18,9,46]. A 2010 critique [2] attracted particular attention and provoked a variety of responses [19,3,20,21]. Here we examine the state of the inclusive fitness controversy six years after this paper. We point out how, despite the appearance of unanimity, the responses of the inclusive fitness community disagree on central issues. Within this discussion, we identify two points which may suggest a path to future agreement: first, there is no consistent way to define the inclusive fitness of an individual [8\*\*,22], and second, evolution does not, in general, lead individuals to maximize their inclusive fitness [7\*\*,11\*\*].

These two points, which have emerged from careful mathematical analysis, contradict what have long been seen as central principles of inclusive fitness theory. It remains to be seen how the larger inclusive fitness community will respond to these points, and whether they can form the basis of an eventual resolution of the controversy.

### Conflicting definitions of inclusive fitness

A letter with over one hundred signatories, written in response to Nowak et al. [2], stated that 'Inclusive fitness is as general as the genetical theory of natural selection itself' [3]. This letter outwardly suggests that the inclusive fitness community has responded with a single voice to defend a coherent and unproblematic theory. However, the various responses to [2] define inclusive fitness in different and sometimes contradictory ways [20,21,23,24,25\*\*]. As a recent paper puts it, 'there is a consensus that inclusive fitness 'works', but none regarding how to explicitly measure and test it' [26].

First, some argue that the limitations of inclusive fitness are not overly limiting [27,24,28\*\*], because the needed additivity follows from assuming that mutations have phenotypically small effect. To illustrate this argument, consider two competing genetic types that are close to each other — say  $g_i = 0$  for the wild-type and  $g_i = \epsilon \ll 1$  for the mutant — on some relevant numerical scale. Assuming that the fitness of each individual  $i$  is a differentiable

function of the form  $w_i(g_1, \dots, g_N)$ , we have the Taylor expansion

$$w_i(g_1, \dots, g_N) = w_i(0, \dots, 0) + \sum_{j=1}^N g_j \frac{\partial w_i}{\partial g_j}(0, \dots, 0) + \mathcal{O}(\epsilon^2). \quad (4)$$

Thus the additivity requirement (1) holds approximately in this case [29,27,24]. However, while the assumptions of differentiability and small phenotypic differences are mathematically convenient, and are invoked in other contexts such as adaptive dynamics [30–34], they are not the general case of evolution. Single mutations can have dramatic phenotypic effects [35]. Moreover, these assumptions are not needed in other theoretical approaches [36,37,38\*\*,39\*\*,40\*].

Second, some deny that inclusive fitness has any limitations at all [3,20,41,42]. This argument is based on a method that determines the 'cost' and 'benefit' of a trait retrospectively, from a predetermined outcome. Given the fitness  $w_i$ , genotype  $g_i$ , and average partners' genotype  $g'_i$  for each individual  $i$ , one proceeds as if fitting a least-squares linear regression

$$w_i = \text{intercept} + \beta_{w,g|g'} g_i + \beta_{w,g'|g} g'_i + \epsilon_i. \quad (5)$$

Simple algebra [43,20,17,41,42] shows that the direction of genetic change agrees with the sign of  $\beta_{w,g|g'} + \beta_{w,g'|g} R$ , where  $R = \text{Cov}[g, g'] / \text{Var}[g]$  is interpreted as relatedness. It is then assumed — incorrectly — that the partial regression coefficients equal the actual fitness effects caused by an individual's genotype, with  $\beta_{w,g'|g}$  representing the benefit to others and  $-\beta_{w,g|g'}$  the cost to self. This method is flawed in numerous ways. First, since the outcome of selection (i.e. the fitnesses  $w_i$ ) must be specified before the 'benefit'  $\beta_{w,g'|g}$  and 'cost'  $-\beta_{w,g|g'}$  are determined, this method cannot yield any predictions or testable hypotheses [47,17]. Second, the regression coefficients do not capture causal fitness effects; they disagree with straightforward calculations of the fitness changes caused by a hypothetical intervention [44\*,45,46]. Third, this method produces nonsensical results; for example, a nurse who helps the weakest is mislabeled as 'spiteful', since interacting with the nurse is negatively correlated with fitness [17,25\*\*]. Finally, this method does not define or identify the inclusive fitness of any individual; rather, the regression coefficients  $\beta_{w,g|g'}$  and  $\beta_{w,g'|g}$  depend on the entire population process (including relatedness).

Third, some have worked to extend inclusive fitness theory beyond the limitation of additive fitness effects [48,28\*\*,49\*]. These approaches involve adding nonlinear, 'synergistic' terms to Eqn 1 for fitness. An effect on  $i$ 's fitness due to the synergistic action of individuals  $j$  and  $k$  can be represented by the term  $\ell_{j,k,i} g_j g_k$  [50,49\*]. In

general, considering interactions among  $m$  individuals, one can write [28\*\*]

$$w_i = w_0 + \sum_{m=1}^N \sum_{j_1, \dots, j_m} e_{j_1, \dots, j_m; i} g_{j_1} \dots g_{j_m}. \quad (6)$$

Under additional assumptions of weak selection and fixed population structure, the products  $g_{j_1} \dots g_{j_m}$  can be replaced by appropriately defined multi-relatedness coefficients  $r_{j_1, \dots, j_m}$ , yielding

$$w_i = w_0 + \sum_{m=1}^N \sum_{j_1, \dots, j_m} e_{j_1, \dots, j_m; i} r_{j_1, \dots, j_m}. \quad (7)$$

This approach, the mathematics of which are still being developed, provides a potentially rigorous way to study nonlinear social interactions among relatives. However, Eqn 7 expresses direct fitness, not inclusive fitness. An inclusive fitness analogue would require splitting these synergistic effects into portions caused by particular individuals, which would violate the very concept of synergy. Additionally, this approach leads to combinatorial explosion: for a general interaction in a population of size  $N$  one needs  $2^N$  multi-relatedness coefficients, with a particular fitness effect assigned to each.

Fourth, Bourke [21] has argued that inclusive fitness theory refers not to a particular method based on the quantity of inclusive fitness (2), but to the general concept that ‘selection of genes for social behaviour depends on social effects on genetic co-bearers’ [21]. But this concept has never been disputed [51,52]. The debate is over the value of the quantity ‘inclusive fitness’ for studying this concept.

Finally, some [7\*\*,11\*\*,8\*\*,22,45,9] have concluded that long-held beliefs about inclusive fitness theory are untenable. These works show that, in general, there is no inclusive fitness of an individual, and individuals do not evolve to maximize inclusive fitness. Although they may draw different verbal conclusions, there is no mathematical disagreement between these works and the critiques of Nowak et al. [2] and others [13–16,2,17,18,46]. Any resolution of the inclusive fitness controversy must accept these points as mathematical reality. In the following sections, we show how these two points undermine the primary arguments for the usefulness of inclusive fitness theory.

### There is no inclusive fitness of an individual

The notion that inclusive fitness ‘belongs’ to the individual — in that it depends only on an individual’s own actions and not on the actions of others — has long been claimed as a major advantage of the approach [1,20,53,42]. The idea is that one can explain the evolution of social behavior, without reference to its genetic basis or evolutionary dynamics, solely in terms of how an individual behaves with regard to its relatives.

However, some supporters of inclusive fitness [8\*\*,22] and outside observers [9] now appear to agree with the critics [13,15,2,18] that, in general, there is no inclusive fitness of an individual. In an article entitled ‘There is no fitness but fitness, and the lineage is its bearer’ [8\*\*], Akçay and Van Cleve argue that inclusive fitness, defined properly, is a property not of the individual but of the *genetic lineage*, which consists of all gene copies descended from a single initial mutation. Furthermore, at the level of the genetic lineage, there is no distinction between direct and inclusive fitness: ‘Inclusive fitness is simply classical Darwinian fitness, averaged over social, environmental and demographic states that members of a gene lineage experience. Therefore, inclusive fitness is neither a generalization of classical fitness, nor does it belong exclusively to the individual’ [8\*\*]. In this view, direct fitness and inclusive fitness are two different ways of expressing a lineage’s *invasion fitness*, defined as its growth rate while rare [54,22].

The non-existence of individual-level inclusive fitness is not a semantic problem but a mathematical one [2,9]. In order to formulate the inclusive fitness of an individual as in Eqn 2, one must be able to identify the particular fitness effect that this individual has on the fitnesses of all others. But since nonlinear, synergistic phenomena cannot be attributed to individual actors, there is in general no meaningful way to define an individual’s inclusive fitness (Figure 2). The regression method [43,20,41,42] does not provide escape from this conundrum, since the partial regression coefficients in Eqn 5 do not describe the actual fitness effects caused by an individual [17,45,44\*]. Moreover, these coefficients apply to the whole population and not to any particular individual. Neither does the multi-relatedness approach [48,28\*\*,49\*] resolve this problem, since ‘credit’ for synergistic fitness effects goes to subsets (pairs, triples, etc.), not to individuals.

Akçay and Van Cleve [8\*\*] propose to define inclusive fitness at the level of the genetic lineage rather than the individual. However, the idea that inclusive fitness is under an individual’s control has long been cited as a major advantage of the theory: ‘No other definition of fitness provides a measure that is both a target of selection and also under the full control of the organism’ [53]. It must now be recognized that this argument rests on a false premise.

### Individuals do not maximize inclusive fitness

Inclusive fitness theory is also said to explain the appearance of design in nature, in that ‘organisms should appear designed to maximise their inclusive fitness’ [53]. Again, the mathematics of natural selection shows that this idea is valid only in special cases [44\*]. If fitness effects are not additive, there is no inclusive fitness of an individual, and



therefore no quantity to maximize (Figure 2). Furthermore, it is well-known that natural selection can lead to cyclic or chaotic behavior [55], precluding the existence of maximands.

Claims that natural selection leads to inclusive-fitness maximization cite Grafen [56], but his approach assumes that fitness effects are additive, and has both mathematical and conceptual shortcomings [57–59]. Another argument [60], based on the regression method, is simply incorrect: it assumes — falsely [17,25<sup>\*\*</sup>,9,45,44<sup>\*</sup>,46] — that the partial regression coefficients in Eqn 5 are actual amounts of fitness under an individual's control.

An interesting question is whether there is *any* quantity that individuals evolve to act ‘as if’ maximizing. Several recent papers [10,7<sup>\*\*</sup>,11<sup>\*\*</sup>,12] have investigated this question by looking for a utility function  $U$  such that, at evolutionarily stable states, individuals behave as if maximizing  $U$ . Lehmann et al. [7<sup>\*\*</sup>] study a metapopulation model and find that in general, no such  $U$  exists. If one assumes weak selection, a maximand  $U$  exists, but it is not inclusive fitness, except under the additional assumption of additive fitness effects. In the context of game-theoretical models with an assortment parameter  $r$ , Alger and Weibull [10] and Okasha and Martens [11<sup>\*\*</sup>] find that selection leads individuals to act as if maximizing the quantity

$$(1-r) \times (\text{payoff vs. opponent's strategy}) \\ + r \times (\text{payoff vs. own strategy}).$$

Again, this quantity is not ‘inclusive fitness’ by any established definition; it coincides with inclusive fitness only when the game has additive payoffs.

Therefore natural selection does not, in general, lead individuals to act as if maximizing inclusive fitness. Arguments based on inclusive-fitness maximization can be valid in the special case that fitness effects are additive, but otherwise have heuristic value at best.

### Kin selection

Maynard Smith [61] coined the term ‘kin selection’ to describe selection of traits via their effect on close relatives. The mechanism of kin selection is indirect replication — genes affecting the fitness of copies of themselves in relatives. In the current literature, however, the term ‘kin selection’ is used so expansively that it applies even when individuals are related neither by coancestry nor genotype [20]. We prefer to restrict ‘kin selection’ to behaviors directed at specific relatives (siblings, nephews, etc.).

The question remains, how should kin selection be modeled mathematically? As recent work [2,17,10,

18,7<sup>\*\*</sup>,25<sup>\*\*</sup>,11<sup>\*\*</sup>,12,8<sup>\*\*</sup>,44<sup>\*</sup>,45,9,46] has made clear, the traditional, individual-centered formulation of inclusive fitness is not the fully general tool it is often claimed to be. We suggest instead to model kin selection at the level of genes, using population genetics and evolutionary dynamics.

This straightforward, gene-centered approach opens the door to new insights. Take, for example, the evolution of eusocial behavior in insects. It is often claimed, based on inclusive fitness reasoning [62–64], that queen monogamy favors both the origin of eusociality and its advancement to a state where workers do not reproduce. However, the inclusive fitness arguments for this idea implicitly assume weak selection and additive fitness effects, without any particular biological justification. Simulations [65–67] and mathematical analyses [39<sup>\*\*</sup>] based on population genetics, which do not make such assumptions, have instead found mixed results as to whether monogamy favors the origin of eusociality and the emergence of non-reproducing worker castes, depending in both cases on ecological details. These results suggest that many questions in the evolution of social behavior — including those commonly seen as ‘settled’ by inclusive fitness theory — stand to benefit from a rigorous analysis based on gene-level evolutionary dynamics [39<sup>\*\*</sup>,40<sup>\*</sup>].

### Conclusions

After years of debate over the usefulness and generality of inclusive fitness [13,15,2,3,20,17,25<sup>\*\*</sup>,46], there appears to be emerging consensus on two key mathematical facts. Recent papers by supporters of inclusive fitness theory [8<sup>\*\*</sup>,7<sup>\*\*</sup>,22] and outside observers [11<sup>\*\*</sup>,9,45,44<sup>\*</sup>] agree with our critique on two points: for general evolutionary processes (i) there is no inclusive fitness of an individual, and (ii) natural selection does not cause individuals to act as if maximizing their inclusive fitness.

The study of social evolution can now move forward on the basis of these facts. Some may still feel that inclusive fitness has a conceptual appeal that transcends its mathematical limitations. We argue, however, that the gene-centered evolutionary dynamics framework is mathematically and conceptually complete, and allows for meaningful and rigorous investigation of the evolution of social behavior — including kin selection — without evoking the problematic quantity of ‘inclusive fitness’.

### Conflict of interest statement

Nothing declared.

### Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Hamilton WD: **The genetical evolution of social behaviour. I.** *J Theor Biol* 1964, **7**:1-16.
2. Nowak MA, Tarnita CE, Wilson EO: **The evolution of eusociality.** *Nature* 2010, **466**:1057-1062.
3. Abbot P, Abe J, Alcock J, Alizon S, Alpedrinha JA, Andersson M, Andre JB, van Baalen M, Balloux F, Balshine S et al.: **Inclusive fitness theory and eusociality.** *Nature* 2011, **471**:E1-E4.
4. Gore J, Youk H, Van Oudenaarden A: **Snowdrift game dynamics and facultative cheating in yeast.** *Nature* 2009, **459**:253-256.
5. Archetti M, Ferraro DA, Cristofori G: **Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer.** *Proc Natl Acad Sci U S A* 2015, **112**:1833-1838.
6. Kelsic ED, Zhao J, Vetsigian K, Kishony R: **Counteraction of antibiotic production and degradation stabilizes microbial communities.** *Nature* 2015, **521**:519.
7. Lehmann L, Alger I, Weibull J: **Does evolution lead to maximizing behavior?** *Evolution* 2015, **69**:1858-1873.  
 Uses a metapopulation model to show that, in general, natural selection does not lead individuals to act as if maximizing any quantity. A maximand exists under weak selection, but this maximand is not inclusive fitness unless fitness effects are additive.
8. Akçay E, Van Cleve J: **There is no fitness but fitness, and the lineage is its bearer.** *Philos Trans R Soc Lond B: Biol Sci* 2016, **371**:20150085.  
 Argues that inclusive fitness is not a property of the individual but of the genetic lineage, that is, the descendants of a single mutant gene. Moreover, inclusive fitness and direct fitness not distinct; they are both equivalent to invasion fitness.
9. Birch J: **Hamilton's two conceptions of social fitness.** *Philos Sci* 2016. <http://www.journals.uchicago.edu/doi/abs/10.1086/687869> (in press).
10. Alger I, Weibull JW: **Homo moralis – preference evolution under incomplete information and assortative matching.** *Econometrica* 2013, **81**:2269-2302.
11. Okasha S, Martens J: **Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games.** *J Evol Biol* 2015, **29**:473-482.  
 Shows that, under replicator dynamics with assortment, selection leads individuals maximize a quantity that is not inclusive fitness. Inclusive fitness only becomes a maximand when fitness effects are additive.
12. Alger I, Weibull JW: **Evolution and Kantian morality.** *Games Econ Behav* 2016, **98**:56-67.
13. Cavalli-Sforza LL, Feldman MW: **Darwinian selection and "altruism".** *Theor Popul Biol* 1978, **14**:268-280.
14. Charlesworth B: **Some models of the evolution of altruistic behaviour between siblings.** *J Theor Biol* 1978, **72**:297-319.
15. Karlin S, Matessi C: **The eleventh RA Fisher memorial lecture: kin selection and altruism.** *Proc R Soc Lond B: Biol Sci* 1983, **219**:327-353.
16. Matessi C, Karlin S: **On the evolution of altruism by kin selection.** *Proc Natl Acad Sci U S A* 1984, **81**:1754-1758.
17. Allen B, Nowak MA, Wilson EO: **Limitations of inclusive fitness.** *Proc Natl Acad Sci U S A* 2013, **110**:20135-20139.
18. van Veelen M, Luo S, Simon B: **A simple model of group selection that cannot be analyzed with inclusive fitness.** *J Theor Biol* 2014, **360**:279-289.
19. Krakauer DC, Flack JC: **Better living through physics.** *Nature* 2010, **467**:661.
20. Gardner A, West SA, Wild G: **The genetical theory of kin selection.** *J Evol Biol* 2011, **24**:1020-1043.
21. Bourke AFG: **The validity and value of inclusive fitness theory.** *Proc R Soc B: Biol Sci* 2011, **278**:3313-3320.
22. Lehmann L, Mullon C, Akçay E, Van Cleve J: **Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations.** *Evolution* 2016, **70**:1689-1702.
23. Birch J: **Hamilton's rule and its discontents.** *Br J Philos Sci* 2014, **65**:381-411.
24. Lehmann L, Rousset F: **The genetical theory of social behaviour.** *Philos Trans R Soc B: Biol Sci* 2014, **369**:20130357.
25. Birch J, Okasha S: **Kin selection and its critics.** *BioScience* 2015, **65**:22-32.  
 Clarifies various aspects of the controversy over inclusive fitness theory, including conflicting definitions of Hamilton's rule and the problem of assigning causal meaning to partial regression coefficients.
26. Nonacs P, Richards MH: **How (not) to review papers on inclusive fitness.** *Trends Ecol Evol* 2015, **20**:1-2.
27. Wakano JY, Ohtsuki H, Kobayashi Y: **A mathematical description of the inclusive fitness theory.** *Theor Popul Biol* 2013, **84**:46-55.
28. van Cleve J: **Social evolution and genetic interactions in the short and long term.** *Theor Popul Biol* 2015, **103**:2-26.  
 Presents a mathematical framework for understanding the evolution of social behavior in the short and long term, using coalescent theory and adaptive dynamics.
29. Wild G, Traulsen A: **The different limits of weak selection and the evolutionary dynamics of finite populations.** *J Theor Biol* 2007, **247**:382-390.
30. Nowak M, Sigmund K: **The evolution of stochastic strategies in the prisoner's dilemma.** *Acta Appl Math* 1990, **20**:247-265.
31. Dieckmann U, Law R: **The dynamical theory of coevolution: a derivation from stochastic ecological processes.** *J Math Biol* 1996, **34**:579-612.
32. Metz JAJ, Geritz SAH, Meszéna G, Jacobs FA, van Heerwaarden JS: **Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction.** In *Stochastic and Spatial Structures of Dynamical Systems*. Edited by van Strien SJ, Lunel SMV. Amsterdam: KNAW Verhandelingen, Afd.; 1996:183-231.
33. Doebeli M: *Adaptive Diversification, vol 48 of Monographs in Population Biology*. Princeton University Press; 2011.
34. Allen B, Nowak MA, Dieckmann U: **Adaptive dynamics with interaction structure.** *Am Nat* 2013, **181**:E139-E163.
35. Fiegna F, Yuen-Tsu NY, Kadam SV, Velicer GJ: **Evolution of an obligate social cheater to a superior cooperator.** *Nature* 2006, **441**:310-314.
36. Nowak MA, Sigmund K: **Evolutionary dynamics of biological games.** *Science* 2004, **303**:793-799.
37. Simon B: **Continuous-time models of group selection, and the dynamical insufficiency of kin selection models.** *J Theor Biol* 2014, **349**:22-31.
38. Allen B, Nowak MA: **Games among relatives revisited.** *J Theor Biol* 2015, **378**:103-116.  
 Shows how cooperation among siblings in a finite population can be analyzed directly at the level of genes. Inclusive fitness analysis is not possible in the general case of this model, because the inclusive fitness of an individual only exists when fitness effects are additive.
39. Olejarz JW, Allen B, Veller C, Nowak MA: **The evolution of non-reproductive workers in insect colonies with haplodiploid genetics.** *eLife* 2015:e08918.  
 Uses evolutionary dynamics to analyze the evolution of worker sterility in insect colonies. Queen monogamy can either promote or hinder worker sterility, depending on ecological details.
40. Olejarz JW, Allen B, Veller C, Gadagkar R, Nowak MA: **Evolution of worker policing.** *J Theor Biol* 2016, **399**:103-116.

Uses evolutionary dynamics to analyze the evolution of policing behavior among workers in insect colonies.

41. Rousset F: **Regression, least squares, and the general version of inclusive fitness.** *Evolution* 2015, **69**:2963-2970.
42. Marshall JA: *Social Evolution and Inclusive Fitness Theory: An Introduction.* Princeton University Press; 2015.
43. Queller DC: **A general model for kin selection.** *Evolution* 1992, **46**:376-380.
44. Okasha S, Martens J: **The causal meaning of Hamilton's rule.** *R Soc Open Sci* 2016, **3**:160037.  
Shows that, in general, the 'benefit' and 'cost' one obtain from regression-based formulations of Hamilton's rule do not agree with straightforward counterfactual notions of benefit and cost.
45. Okasha S: **On Hamilton's rule and inclusive fitness theory with non-additive payoffs.** *Philos Sci* 2016.
46. van Veelen M, Allen B, Hoffman M, Simon B, Veller C: **Hamilton's rule.** *J Theor Biol* 2016. <http://www.sciencedirect.com/science/article/pii/S0022519316302533> (in press).
47. Chuang JS, Rivoire O, Leibler S: **Cooperation and Hamilton's rule in a simple synthetic microbial system.** *Mol Syst Biol* 2010, **6**.
48. Ohtsuki H: **Evolutionary dynamics of n-player games played by relatives.** *Philos Trans R Soc Lond B: Biol Sci* 2014, **369**:20130359.
49. Taylor P: **Hamilton's rule in finite populations with synergistic interactions.** *J Theor Biol* 2016, **397**:151-157.  
Develops the multi-relatedness approach to the evolution of social behavior.
50. Queller DC: **Kinship, reciprocity and synergism in the evolution of social behaviour.** *Nature* 1985, **318**:366-367.
51. Nowak MA, Tarnita CE, Wilson EO: **Nowak et al. reply.** *Nature* 2011, **471**:E9-E10.
52. Nowak MA, Allen B: **Inclusive fitness theorizing invokes phenomena that are not relevant for the evolution of eusociality.** *PLoS Biol* 2015, **13**:e1002134.
53. West SA, Gardner A: **Adaptation and inclusive fitness.** *Curr Biol* 2013, **23**:R577-R584.
54. Metz J, Nisbet R, Geritz S: **How should we define 'fitness' for general ecological scenarios?** *Trends Ecol Evol* 1992, **7**:198-202.
55. Hofbauer J, Sigmund K: *Evolutionary Games and Replicator Dynamics.* Cambridge, UK: Cambridge University Press; 1998.
56. Grafen A: **Optimization of inclusive fitness.** *J Theor Biol* 2006, **238**:541-563.
57. Ewens WJ: **Grafen, the price equations, fitness maximization, optimisation and the fundamental theorem of natural selection.** *Biol Philos* 2014, **29**:197-205.
58. Lehmann L, Rousset F: **Fitness, inclusive fitness, and optimization.** *Biol Philos* 2014, **29**:181-195.
59. Birch J: **Natural selection and the maximization of fitness.** *Biol Rev* 2015, **91**:712-727.
60. Gardner A, Welch JJ: **A formal theory of the selfish gene.** *J Evol Biol* 2011, **24**:1801-1813.
61. Smith JM: **Group selection and kin selection.** *Nature* 1964, **201**:1145-1147.
62. Boomsma JJ: **Lifetime monogamy and the evolution of eusociality.** *Philos Trans R Soc Lond B: Biol Sci* 2009, **364**:3191-3207.
63. Gardner A, Alpedrinha J, West SA: **Haplodiploidy and the evolution of eusociality: split sex ratios.** *Am Nat* 2012, **179**:240-256.
64. Alpedrinha J, Gardner A, West SA: **Haplodiploidy and the evolution of eusociality: worker revolution.** *Am Nat* 2014, **184**:303-317.
65. Nonacs P: **Monogamy and high relatedness do not preferentially favor the evolution of cooperation.** *BMC Evol Biol* 2011, **11**:1.
66. Fromhage L, Kokko H: **Monogamy and haplodiploidy act in synergy to promote the evolution of eusociality.** *Nat Commun* 2011, **2**:397.
67. Nonacs P: **Resolving the evolution of sterile worker castes: a window on the advantages and disadvantages of monogamy.** *Biol Lett* 2014, **10**:20140089.