Inclusive Fitness Maximization: An Axiomatic Approach

Samir Okasha[a*], John A. Weymark[b], Walter Bossert[c]

[a] *Department of Philosophy, University of Bristol, 9 Woodland Road, Bristol BS8 1TB, United Kingdom*
[b] *Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA*
[c] *Department of Economics and CIREQ, University of Montreal, P.O. Box 6128, Station Downtown, Montreal QC H3C 3J7, Canada*

May 2013

Abstract. Kin selection theorists argue that evolution in social contexts will lead organisms to behave as if maximizing their inclusive, as opposed to personal, fitness. The inclusive fitness concept allows biologists to treat organisms as akin to rational agents seeking to maximize a utility function. Here we develop this idea and place it on a firm footing by employing a standard decision-theoretic methodology. We show how the principle of inclusive fitness maximization and a related principle of quasi-inclusive fitness maximization can be derived from axioms on an individual's 'as if preferences' (binary choices). Our results help integrate evolutionary theory and rational choice theory, help draw out the behavioural implications of inclusive fitness maximization, and point to a possible way in which evolution could lead organisms to implement it.

Keywords. Hamilton's Rule, inclusive fitness, kin selection, rational choice

[*] Corresponding author. Tel. +44 (0) 117 928 7829
E-mail addresses: samir.okasha@bristol.ac.uk, john.weymark@vanderbilt.edu, walter.bossert@umontreal.ca

1. Introduction

A central tenet of inclusive fitness theory is that a trait may be selected for even if it involves some sacrifice to personal fitness, provided that it sufficiently enhances the reproductive success of genetically related individuals. Typically, genetic relatedness between social partners stems from kinship, in which case inclusive fitness theory can be identified with kin selection theory. Inclusive fitness is central to much work on the evolution of social behaviour. It has been used to understand diverse biological phenomena including sex-ratios, co-operative breeding, dispersal, reproductive skew, group formation, and more. For introductions to inclusive fitness theory, see Frank (1998), McElreath and Boyd (2007), and Wenseleers, Gardner, and Foster (2010).

J. B. S. Haldane purportedly enunciated the basic idea of inclusive fitness theory in a pub when he quipped that he would sacrifice himself by jumping into a river to save two brothers or eight cousins, a view he only expressed in print at a much later date (see Haldane (1955, p. 44)). However, it was W. D. Hamilton (1963, 1964a,b) who first provided a precise formal statement of the theory. In addition to Haldane (1955), other precursors to Hamilton include Darwin (1859), Fisher (1930), and Haldane (1932) (see Dugatkin (2007)).

Hamilton's original theory contains two distinct though related ideas: firstly, his famous rule for when a gene coding for an altruistic action will be favoured by natural selection; and secondly the idea of inclusive fitness, as opposed to personal fitness, as the quantity that individuals will behave as if they are trying to maximize. *Hamilton's Rule* is expressed by the inequality $rb > c$. This rule tells us that a gene for altruism will spread so long as the cost c to the altruist is offset by a sufficient amount of benefit b to relatives who are sufficiently close, as measured by the relatedness coefficient r. This way of thinking involves taking the 'gene's eye view', that is, looking for the selective advantage that a trait has for the gene that causes the trait, rather than the individual that expresses it. However, Hamilton showed that altruistic behaviour can also be understood from an individual's perspective. Though an individual performing an altruistic action will reduce its personal fitness (i.e., expected number of offspring), it may enhance its *inclusive* fitness—a measure that also takes into account the effect of the action on the reproductive output of relatives. Under certain conditions, it can be shown that natural selection will lead an individual to behave as if it is trying to maximize its inclusive fitness (see Frank, 1998; McElreath and

Boyd, 2007; Grafen, 2006, 2009).

The concept of inclusive fitness is somewhat unintuitive, and critics have questioned both the generality of the theory and the usefulness of the concept (e.g., Nowak, Tarnita, and Wilson, 2010). While granting that inclusive fitness has its limitations, and that there are other valid ways to study the evolution of social behaviour, here we focus on a conceptually attractive feature of inclusive fitness theory, namely that it allows us to preserve the idea of the individual organism as a quasi-rational agent, choosing between alternative actions according to the criterion of *maximal inclusive fitness*. This aspect of the theory explains its wide appeal to behavioural ecologists as it allows them to take an adaptationist approach to social behaviour, as has been emphasized in recent work by Grafen (2006, 2009) and Gardner, West, and Wild (2011), among others.

In this article, we offer a novel perspective on inclusive fitness theory by applying tools from the economic theory of rational choice. Our aim is to derive inclusive fitness maximization from axioms on an individual organism's choice behaviour. Consider a focal individual and the set of other individuals who might be affected by this individual's actions. Each of the latter individuals stands in a fixed relatedness relationship to the focal individual. The focal individual is faced with a choice between alternative social actions. Each action leads to a payoff (which could be positive, negative, or zero) for the focal individual and each of the other affected individuals. An individual's payoff is the incremental change in its personal fitness due to the focal individual's action. The focal individual's choice behaviour is described by a binary preference relation on the set of actions. This relation specifies, for any two actions, which the focal individual would choose; in principle, this choice could be directly observed. The question we pose is: What conditions must this binary relation satisfy such that the focal individual always behaves as if it were trying to maximize its inclusive fitness? We also consider a variant of inclusive fitness maximization called *quasi-inclusive fitness maximization* that can be applied when the focal individual is unable to determine the exact degree of relatedness to some of the other individuals, and axiomatically characterize this behaviour as well.

The axiomatic approach employed here is the standard way of justifying a maximization assumption in rational choice theory, and it is instructive to apply it to inclusive fitness for three reasons. Firstly, it offers a novel way of forging links, both formal and conceptual, between social evolution theory and economic theory. Many authors have drawn attention to the analogy between the utility-maximizing paradigm of economics and the fitness-maximizing paradigm of behavioural ecology; here we develop this analogy in a precise way by finding the behavioural conditions that are necessary and sufficient for an organism to be representable as an inclusive fitness maximizer. Our results draw on related work in social choice theory, which is the branch of rational choice theory that is concerned with social preferences. Axiomatic social choice theory has been used by Okasha (2009) and Bossert, Qi, and Weymark (2013a,b) to evaluate alternative measures of group fitness in hierarchically structured populations. This article is the first to apply this methodology to analyzing inclusive fitness.

Secondly, our results suggest a possible route by which evolution could program organisms to implement inclusive fitness maximization, or something close to it. That is, the axioms we use to characterize inclusive fitness maximization could be viewed as heuristic rules by which evolution might induce organisms to display optimal behaviour in social settings without having to consciously perform inclusive fitness calculations.

Thirdly, our results help bring out the behavioural implications of inclusive fitness theory, and could thus facilitate its empirical testing. An organism's binary choices between actions can be directly observed, whereas the consequences of those choices for inclusive fitness are typically difficult to determine. If it could be shown that an organism's choice behaviour violated one of the axioms below, we could immediately infer that the organism was not maximizing inclusive fitness.

Section 2 describes the formal framework employed here. Our axioms are introduced in Section 3. Our axiomatic characterizations of the two forms of inclusive fitness maximization are presented in Section 4. We offer some concluding remarks in Section 5. The proofs of our theorems may be found in the Appendix.

2. The Model

We consider a set of individuals $I = \{1, \ldots, n\}$. Individual 1 is the focal individual whose actions we are interested in; the other $n - 1$ comprise all the other individuals who might be affected by the focal individual's actions. We let $r_i \in [0, 1]$ denote the relatedness of the focal individual to individual i, with higher values denoting a closer degree of relatedness; so $r_1 = 1$ and $r_i \leq 1$ for all $i \neq 1$. Thus, the set I has an associated relatedness profile $\mathbf{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$, where $\mathbf{0}$ denotes an n-vector of zeros. The profile $\mathbf{r}$ is a fixed parameter of our model. We set aside the much-debated question of how exactly relatedness should be defined in inclusive fitness theory; our formal model is neutral with respect to this, requiring only that all relatednesses satisfy the restrictions described above.

The focal individual can perform a number of different actions, each of which potentially affects the personal fitness (expected number of offspring) of every individual in I. We identify an action with a payoff vector $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$, where $a_i \in \mathbb{R}$ is the incremental personal fitness gain or loss that individual i suffers as a result of action $\mathbf{a}$. The set of all possible actions is $\mathbb{R}^n$. There is a fixed status-quo payoff vector $\mathbf{s} = (s_1, \ldots, s_n) \in \mathbb{R}^n_+$ describing the fitness of each individual before any action is performed. Thus, the set of feasible actions is given by $F = \{\mathbf{a} \in \mathbb{R}^n \mid \mathbf{a} + \mathbf{s} \geq \mathbf{0}\} = \{\mathbf{a} \in \mathbb{R}^n \mid \mathbf{a} \geq -\mathbf{s}\}$.

The focal individual's choice behaviour is described by a binary preference relation $\succsim_{\mathbf{r}}$ on F. The relation $\succsim_{\mathbf{r}}$ indicates, for any two actions in F, which the focal individual would prefer given the relatedness profile $\mathbf{r}$; formally, $\succsim_{\mathbf{r}}$ is a subset of $F \times F$. Although the relatedness profile $\mathbf{r}$ is fixed here, we include it in the notation for this binary relation to highlight its conditionality on $\mathbf{r}$. As the notation suggests, $\succsim_{\mathbf{r}}$ is a weak preference relation; that is, $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ means that action $\mathbf{a}$ is either strictly preferred or indifferent to $\mathbf{b}$. From $\succsim_{\mathbf{r}}$, we can define corresponding relations of strict preference $\succ_{\mathbf{r}}$ and of indifference $\sim_{\mathbf{r}}$ by letting $\mathbf{a} \succ_{\mathbf{r}} \mathbf{b} \equiv_{\mathit{df}} [\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ and not$(\mathbf{b} \succsim_{\mathbf{r}} \mathbf{a})]$ and $\mathbf{a} \sim_{\mathbf{r}} \mathbf{b} \equiv_{\mathit{df}} [\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ and $\mathbf{b} \succsim_{\mathbf{r}} \mathbf{a}]$. The concept of preference being appealed to here is an 'as if' one; the preference $\succsim_{\mathbf{r}}$ is simply a way of summarizing the focal individual's choice behaviour. That is, $\mathbf{a} \succ_{\mathbf{r}} \mathbf{b}$ means that $\mathbf{a}$ is chosen when the options are $\mathbf{a}$ and $\mathbf{b}$, whereas $\mathbf{a} \sim_{\mathbf{r}} \mathbf{b}$ means that either of these actions might be chosen when both are available.

The **inclusive fitness** of a feasible action $\mathbf{a} \in F$ is defined as $\sum_{i=1}^n r_i a_i$. That is, it is a weighted sum over individuals of the action's payoff to each individual, with weights given by the relatedness profile. If the focal individual is an **inclusive fitness maximizer**, then its preference relation $\succsim_{\mathbf{r}}$ is represented by the inclusive fitness function, which means that for all actions $\mathbf{a}, \mathbf{b} \in F$, $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ if and only if $\sum_{i=1}^n r_i a_i \geq \sum_{i=1}^n r_i b_i$.

If the focal individual is not an inclusive fitness maximizer, this may be because it cannot discriminate sufficiently precisely between different classes of relatives. We define a **quasi-inclusive fitness maximizer** as an individual whose preference relation $\succsim_{\mathbf{r}}$ is represented by $\sum_{i=1}^n \beta_i a_i$ for some vector $(\beta_1, \ldots, \beta_n) \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ such that $\beta_i > \beta_j$ if and only if $r_i > r_j$ for all $i, j \in I$. A quasi-inclusive fitness maximizer uses a weighted sum of the payoffs to evaluate an action; however, the weights need not be the true relatednesses but, rather, can be any monotonic transformation of them.

The concept of quasi-inclusive fitness maximization is interesting for two different reasons. Firstly, it describes a way that an organism might attempt to maximize inclusive fitness if it lacks information about exact degrees of relatedness, but can tell who it is more related to. Empirically, it seems likely that many organisms are in this situation. Secondly, it highlights the fact that inclusive fitness maximization comprises two logically separate components: (i) evaluating social actions by a weighted sum of the payoffs and (ii) using relatednesses as the weights in the sum. Below, we obtain an axiomatic separation of these two components of inclusive fitness theory.

Our goal is to identify 'natural' axioms on $\succsim_{\mathbf{r}}$ that characterize the focal individual as an inclusive fitness maximizer and as a quasi-inclusive fitness maximizer.

We have implicitly assumed that the payoffs (i.e., the incremental fitnesses) are measurable on an absolute scale. This is a stronger assumption than is necessary; both inclusive fitness maximization and quasi-inclusive fitness maximization only require that gains and losses of incremental fitness are comparable across individuals. The importance of measurement-theoretic issues for the quantification of fitness has recently been stressed by Wagner (2010).

3. The Axioms

In this section, we consider a number of axioms that might be imposed on the relation $\succsim_{\mathbf{r}}$ and comment briefly on their meaning and biological significance.

The binary relation $\succsim_{\mathbf{r}}$ is (i) *reflexive* if for all $\mathbf{a} \in F$, $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{a}$, (ii) *complete* if for all $\mathbf{a}, \mathbf{b} \in F$ with $\mathbf{a} \neq \mathbf{b}$, $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ or $\mathbf{b} \succsim_{\mathbf{r}} \mathbf{a}$, and (iii) *transitive* if for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in F$, $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}$ and $\mathbf{b} \succsim_{\mathbf{r}} \mathbf{c}$ imply $\mathbf{a} \succsim_{\mathbf{r}} \mathbf{c}$. An *ordering* is a reflexive, complete, and transitive binary relation.

Ordering. $\succsim_{\mathbf{r}}$ is an ordering.

Ordering is a standard axiom in the theory of rational choice. Essentially it requires that the focal individual can rank all feasible actions in terms of betterness, with ties permitted. Though violations of transitivity have been reported empirically in both humans and animals, this axiom is a fundamental part of the meaning of 'rationality', and is necessary if an individual's choices are to maximize any quantity, inclusive fitness or some other. The reader can easily verify that if the focal individual's choice behaviour violates Ordering, then it is not an inclusive fitness maximizer.

The binary relation $\succsim_{\mathbf{r}}$ is *continuous* if for any action $\mathbf{a} \in F$, the sets $\{\mathbf{b} \in F \mid \mathbf{b} \succsim_{\mathbf{r}} \mathbf{a}\}$ and $\{\mathbf{b} \in F \mid \mathbf{a} \succsim_{\mathbf{r}} \mathbf{b}\}$ are both closed.

Continuity. $\succsim_{\mathbf{r}}$ is continuous.

Continuity is also a standard axiom of rational choice theory. It formalizes the intuitive idea that 'small' changes in payoffs should not lead to 'large' changes in preference. It is an appropriate assumption in any context where payoffs cannot be measured with perfect accuracy or are subject to minor chance fluctuations.

Payoff Dominance. For all $\mathbf{a}, \mathbf{b} \in F$ such that $a_j > b_j$ for all $j \in I$, $\mathbf{a} \succ_{\mathbf{r}} \mathbf{b}$.

Payoff Dominance says that if one action yields a strictly higher payoff for every individual than another action, then the former action is strictly preferred. If the focal individual violated this axiom by choosing a dominated action, then its behaviour would seem clearly non-optimal because by simply switching actions, it would be able to increase the personal fitness of every individual in I. This axiom is closely related to the 'Pareto principle' in social choice theory.

Focal Individual Monotonicity. For all $\mathbf{a}, \mathbf{b} \in F$ such that $a_1 > b_1$ and $a_j = b_j$ for all $j \in \{2, \ldots, n\}$, $\mathbf{a} \succ_{\mathbf{r}} \mathbf{b}$.

Focal Individual Monotonicity says that starting from any action, if the focal individual's payoff is increased while the payoff of all other individuals is held fixed, then the resulting action is strictly preferred to the original. Thus, the focal individual is not *completely* other-regarding; it does care about its own personal fitness. Again, violating this axiom would seem clearly non-optimal for it would amount to sacrificing one's own personal fitness without a compensating gain in personal fitness for anyone else.

Baseline Independence. For all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in F$ such that $(\mathbf{a} + \mathbf{c}) \in F$ and $(\mathbf{b} + \mathbf{c}) \in F$,

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \ \Leftrightarrow\ (\mathbf{a} + \mathbf{c}) \succeq_{\mathbf{r}} (\mathbf{b} + \mathbf{c}).$$

Baseline Independence requires the focal individual's evaluation of an action to be independent of the 'baseline fitnesses' from which we start; so if action $\mathbf{a}$ is preferred to $\mathbf{b}$, this preference will never be reversed by changing the baseline. (Note that on the LHS of the above equivalence, the baseline is the null action $\mathbf{0}$, whereas on the RHS it is $\mathbf{c}$.) So if an individual prefers $\mathbf{a}$ to $\mathbf{b}$ today, it should continue to do so tomorrow, irrespective of what fitness-affecting events have occurred in the interim. Another interpretation is to think of $(\mathbf{b}+\mathbf{c})$ as the result of performing actions $\mathbf{b}$ and $\mathbf{c}$ in succession; the axiom then says that if one action is preferred to another, it should remain so irrespective of which other actions have already been performed.

Nepotism. For all $\mathbf{a}, \mathbf{b} \in F$, for all $j, k \in I$ such that $r_j \geq r_k$, and for all $x > 0$, if $b_j = a_j + x$, $b_k = a_k - x$, and $b_i = a_i$ for all $i \in I \setminus \{j, k\}$, then (i) $\mathbf{b} \succ_{\mathbf{r}} \mathbf{a}$ if $r_j > r_k$ and (ii) $\mathbf{b} \sim_{\mathbf{r}} \mathbf{a}$ if $r_j = r_k$.

Nepotism captures the idea that the focal individual would prefer to help closer than more distant relatives; this is a central prediction of kin selection theory. The axiom says that starting from a given action, if some quantity of payoff is shifted from one individual to another more closely related individual while everyone else's payoff is held fixed, then the resulting action will be preferred; while if payoff is shifted to an equally related individual, indifference will result. To satisfy Nepotism, all the focal individual needs to 'know' is which of any pair of individuals it is more closely related to, but not by how much. This seems a reasonable idealization of the actual powers of kin discrimination of many animals.

Haldane. For all $\mathbf{a}, \mathbf{b} \in F$, if there exist $k \in \{2, \ldots, n\}$ and $x \in \mathbb{R}$ such that (i) $r_k > 0$, $b_1 = a_1 - x$, $b_k = a_k + x/r_k$, and $b_j = a_j$ for all $j \in I \setminus \{1, k\}$ or (ii) $r_k = 0$, $b_1 = a_1$, $b_k = a_k + x$, and $b_j = a_j$ for all $j \in I \setminus \{1, k\}$, then $\mathbf{a} \sim_{\mathbf{r}} \mathbf{b}$.

Haldane provides a formal statement of the idea that starting from a given action, if we reduce the focal individual's own payoff by x and increase the payoff to any other individual i by $\frac{x}{r_i}$, then indifference is the result; that is, the focal individual uses relatedness as the 'exchange rate' for determining which payoff sacrifices it is prepared to make. The axiom derives its name from Haldane's remark quoted in the Introduction that it would be a fitness-enhancing sacrifice to jump into a river to save two brothers or eight cousins when $\mathbf{r} = \left(1, \frac{1}{2}, \frac{1}{8}, \ldots\right)$. Note that this axiom requires only that the focal individual be able to perform 'egocentric' comparisons; that is, it must be able to compare the results of transferring its *own* payoff to others. It does not require comparisons among pairs of actions that involve transfers between two non-focal individuals (unlike Nepotism). Nonetheless, to satisfy Haldane is still a demanding task, as it requires that the focal individual 'knows' its degree of relatedness to every other individual in I, and uses this information to compute the level of self-sacrifice it is prepared to make.

4. The Results

We now use the axioms introduced in the preceding section to provide axiomatic characterizations of inclusive fitness maximization (Theorem 1) and quasi-inclusive fitness maximization (Theorem 2).

Theorem 1. *The relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane if and only if the focal individual is an inclusive fitness maximizer.*

Theorem 1 states necessary and sufficient conditions for the focal individual to be an inclusive fitness maximizer, namely that its preference relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane. It might be thought that this result is somewhat unexciting on the grounds that the Haldane axiom is conceptually quite similar to inclusive fitness maximization itself. However two points should be noted. Firstly, recall that Haldane concerns only 'egocentric' comparisons between actions which involve a transfer of payoff from the focal individual to another. The axiom is silent about how to rank pairs of actions that are not of this sort; yet inclusive fitness maximization yields a ranking of all actions in the feasible set. So the conceptual gap between the axioms of Theorem 1 and the characterization is in fact substantial, and the proof correspondingly non-trivial.

Secondly, note that the Haldane axiom on its own does not suffice to characterize inclusive fitness maximization; the other two axioms of Theorem 1 are also needed. Therefore, the theorem helps to clarify the exact logical relation between Haldane's original idea, as formalized here, and Hamilton's later theory. Because the two axioms that must be added to Haldane to yield inclusive fitness maximization (Ordering and Focal Individual Monotonicity) are fairly obvious rationality requirements, this vindicates the widely-held view that Haldane had grasped the essence of inclusive fitness theory prior its detailed elaboration by Hamilton.

Theorem 2. *The relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Continuity, Payoff Dominance, Baseline Independence, and Nepotism if and only if the focal individual is a quasi-inclusive fitness maximizer.*

Theorem 2 characterizes quasi-inclusive fitness maximization using five axioms that do not include Haldane. As the proof in the Appendix shows, the first four axioms (Ordering, Continuity, Payoff Dominance, and Baseline Independence) imply that the focal individual evaluates actions by a weighted sum of the payoffs for some vector of non-negative weights; the addition of Nepotism then restricts those weights to be monotone transformations of the relatednesses. Thus, the first four axioms characterize one component of inclusive fitness theory—evaluating actions by weighted sums of payoffs, while the fifth axiom ensures a logical link with the second component—using relatednesses as the weights.

Although Theorem 2 only characterizes quasi-inclusive fitness maximization, rather than inclusive fitness maximization itself, it has one significant advantage over Theorem 1, namely, its axioms make weaker informational demands on the focal individual than does Haldane. Consequently, it should be correspondingly easier for natural selection to bring about conformity to them. Recall that Nepotism requires that the focal individual prefers to help closer than more distant relatives; exact degrees of relatedness do not matter. Because kin discrimination is quite common in social species, there is no great difficulty in imagining how natural selection could produce organisms whose choice behaviour satisfies Nepotism. By contrast, it is rather harder to imagine natural selection fine-tuning choice behaviour so as to satisfy Haldane. So although Theorem 2 only yields quasi-inclusive fitness maximization, the axioms it uses are more biologically reasonable.

5. Conclusion

The popularity of the inclusive fitness concept in evolutionary biology arises because it allows social behaviour, even when it is individually costly, to be understood from the perspective of an individual organism ‘trying’ to achieve a goal, thus preserving Darwin’s insight that selection will lead to the appearance of design in nature. (The goal in question, of course, is maximization of inclusive fitness.) This has led many authors to see a link between social evolution and rational choice theory; that is, evolved organisms should behave like rational agents trying to maximize a utility function, where the utility function is inclusive fitness. Our aim has been to develop this idea further and place it on a secure foundation by seeking to deduce inclusive fitness maximization from a more primitive basis, namely axioms on an individual’s ‘as if preferences’, in accordance with standard decision-theoretic methodology. Our hope is that this will shed light on the conceptual links between evolution and rational choice theory, show a possible route by which natural selection could bring about inclusive fitness maximization or something close to it, and help to draw out behavioural implications of inclusive fitness theory that are directly testable.

Appendix

We say that the focal individual is an *m-inclusive fitness maximizer*, $m \in \{2, \dots, n\}$, if, for all $M \subseteq I$ such that $1 \in M$ and $|M| = m$,

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \; \Leftrightarrow \; \sum_{i \in M} r_i a_i \geq \sum_{i \in M} r_i b_i$$

for all $\mathbf{a}, \mathbf{b} \in F$ such that $a_j = b_j$ for all $j \in I \setminus M$. Thus, the focal individual is an inclusive fitness maximizer if it is an n-inclusive fitness maximizer.

The following two lemmas are used in the proof of Theorem 1.

Lemma 1. *If the relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane, then the focal individual is a 2-inclusive fitness maximizer.*

Proof. Consider any $k \in \{2, \dots, n\}$, $M = \{1, k\}$, and $\mathbf{a}, \mathbf{b} \in F$. Let $a'_j = a_j$ for all $j \in I \setminus \{1, k\}$ and consider the set

$$L_{(a_j)_{j \in N \setminus \{1,k\}}}(a_1, a_k) = \{(a'_1, a'_k) \mid \mathbf{a}' \in F \text{ and } \mathbf{a}' \sim_{\mathbf{r}} \mathbf{a}\},$$

where $\mathbf{a}' = (a'_1, \dots, a'_n)$. This is the level set of the restriction of $\succeq_{\mathbf{r}}$ corresponding to the set of components $\{1, k\}$ that contains (a_1, a_k) conditional on the remaining variables having the values $(a_j)_{j \in N \setminus \{1,k\}}$. By subtracting $x = -[r_k(s_k + a_k)]$ from a_1 and adding x/r_k to a_k when $r_k > 0$ or by adding $-(s_k + a_k)$ to a_k when $r_k = 0$, it follows from Haldane that the point $(a_1 + r_k(s_k + a_k), -s_k)$ belongs to this level set.

In order for the focal individual to be a 2-inclusive fitness maximizer, it is necessary that any point (a'_1, a'_k) in $L_{(a_j)_{j \in N \setminus \{1,k\}}}(a_1, a_k)$ be such that

$$a'_1 + r_k a'_k = a_1 + r_k a_k = a_1 + r_k(s_k + a_k) + r_k(-s_k). \tag{1}$$

Any such point can be reached by subtracting $x = a_1 + r_k(s_k + a_k) - a'_1$ from $a_1 + r_k(s_k + a_k)$ and adding x/r_k to $-s_k$ when $r_k > 0$ or by adding $s_k + a'_k$ to $-s_k$ when $r_k = 0$. Thus, by Haldane, it follows that any point

(a'_1, a'_k) for which (1) holds is in the level set of the point $(a_1 + r_k(s_k + a_k), -s_k)$. The transitivity of $\sim_{\mathbf{r}}$ then implies that $\mathbf{a}' \sim_{\mathbf{r}} \mathbf{a}$ for all $(a'_1, a'_k) \in L_{(a_j)_{j \in N \setminus \{1,k\}}}(a_1, a_k)$. By Ordering and Focal Individual Monotonicity, higher level sets of $\succeq_{\mathbf{r}}$ are associated with higher level sets $L_{(a_j)_{j \in N \setminus \{1,k\}}}(a_1, a_k)$.

The same procedure can be applied to $\mathbf{b}$. Defining $\mathbf{b}'$ and $L_{(b_j)_{j \in N \setminus \{1,k\}}}(b_1, b_k)$ by analogy to $\mathbf{a}'$ and $L_{(a_j)_{j \in N \setminus \{1,k\}}}(a_1, a_k)$, it follows that $\mathbf{b}' \sim_{\mathbf{r}} \mathbf{b}$ for all $(b'_1, b'_k) \in L_{(b_j)_{j \in N \setminus \{1,k\}}}(b_1, b_k)$ and that higher level sets of $\succeq_{\mathbf{r}}$ are associated with higher level sets $L_{(b_j)_{j \in N \setminus \{1,k\}}}(b_1, b_k)$. Transitivity now implies that

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; a_1 + r_k a_k \geq b_1 + r_k b_k$$

for all $\mathbf{a}, \mathbf{b} \in F$ such that $a_j = b_j$ for all $j \in I \setminus \{1, k\}$. Hence, the focal individual is a 2-inclusive fitness maximizer. □

The following lemma is established by adapting the proof of Lemma 3.3.1 in d'Aspremont (1985).

Lemma 2. *If the relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane, then the focal individual is an m-inclusive fitness maximizer for all $m \in \{2, \ldots, n\}$.*

Proof. By Lemma 1, the focal individual is a 2-inclusive fitness maximizer. If $n = 2$, we are done. If $n > 2$, we complete the proof by induction. Suppose that the focal individual is an m-inclusive fitness maximizer, where $m \in \{2, \ldots, n-1\}$. We need to show that the focal individual is an $(m+1)$-inclusive fitness maximizer.

It is sufficient to consider the case in which $M = \{1, \ldots, m+1\}$. Let $\mathbf{a}, \mathbf{b} \in F$ be such that $a_j = b_j$ for all $j \in I \setminus \{1, \ldots, m+1\}$. Without loss of generality, we can suppose that $a_{m+1} \geq b_{m+1}$ (if this is not the case, then the roles of $\mathbf{a}$ and $\mathbf{b}$ can be interchanged in the following argument). Define $\mathbf{c} \in \mathbb{R}^n$ by letting

$$c_j = a_j \geq -s_j \quad \forall j \in I \setminus \{1, m+1\}, \tag{2}$$

$$c_{m+1} = b_{m+1} \geq -s_{m+1}, \tag{3}$$

and

$$c_1 = a_1 + r_{m+1}(a_{m+1} - b_{m+1}). \tag{4}$$

Because $a_1 \geq -s_1$ and, by assumption, $a_{m+1} \geq b_{m+1}$, it follows that $c_1 \geq -s_1$ and, together with the inequalities in (2) and (3), we obtain $\mathbf{c} \in F$.

Using (3) and (4), it follows that

$$c_1 + r_{m+1}c_{m+1} = a_1 + r_{m+1}a_{m+1}. \tag{5}$$

By Lemma 1, the focal individual is a 2-inclusive fitness maximizer and, thus, (4) implies

$$\mathbf{c} \sim_{\mathbf{r}} \mathbf{a}. \tag{6}$$

It follows from (2) and (3) that $c_j = b_j$ for all $j \in \{m+1, \ldots, n\}$. By the induction hypothesis, the focal individual is an m-inclusive fitness maximizer and, thus,

$$\mathbf{c} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \sum_{i=1}^{m} r_i c_i \geq \sum_{i=1}^{m} r_i b_i. \tag{7}$$

Because $c_{m+1} = b_{m+1}$, (7) is equivalent to

$$\mathbf{c} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \sum_{i=1}^{m+1} r_i c_i \geq \sum_{i=1}^{m+1} r_i b_i.$$

Furthermore, by (6) and the transitivity of $\succeq_{\mathbf{r}}$,

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \mathbf{c} \succeq_{\mathbf{r}} \mathbf{b}.$$

Thus,

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \sum_{i=1}^{m+1} r_i c_i \geq \sum_{i=1}^{m+1} r_i b_i. \tag{8}$$

Because $c_j = a_j$ for all $j \in I \setminus \{1, m+1\}$ and (5) holds, it follows that

$$\sum_{i=1}^{m+1} r_i a_i = \sum_{i=1}^{m+1} r_i c_i.$$

Substituting this equality in (8), we obtain

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \sum_{i=1}^{m+1} r_i a_i \geq \sum_{i=1}^{m+1} r_i b_i.$$

That is, the focal individual is an $(m+1)$-inclusive fitness maximizer. □

We now use Lemma 2 to prove that the relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane if and only if the focal individual is an inclusive fitness maximizer, as Theorem 1 asserts.

Proof of Theorem 1. It is straightforward to verify that if the focal individual is an inclusive fitness maximizer, then $\succeq_{\mathbf{r}}$ satisfies Ordering, Focal Individual Monotonicity, and Haldane.

Now, suppose that $\succeq_{\mathbf{r}}$ satisfies these three axioms. Lemma 2 states that the focal individual is an m-inclusive fitness maximizer for all $m \in \{2, \dots, n\}$ if $\succeq_{\mathbf{r}}$ satisfies these axioms. Setting $m = n$, it follows that the focal individual is an inclusive fitness maximizer. □

We now turn to the proof of Theorem 2. As a first step, we state a lemma, the proof of which is identical to the proof of Theorem 8.1 in Bossert and Weymark (2004) with a relabeling of the axioms and a change in notation. See also Theorem 4.3.1 in Blackwell and Girshick (1954) for a related result (without the continuity axiom) in the context of decision-making under uncertainty.

Lemma 3. *The relation $\succeq_{\mathbf{r}}$ satisfies Ordering, Continuity, Payoff Dominance, and Baseline Independence if and only if there exists $(\beta_1, \dots, \beta_n) \in \mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ such that, for all $\mathbf{a}, \mathbf{b} \in F$,*

$$\mathbf{a} \succeq_{\mathbf{r}} \mathbf{b} \;\Leftrightarrow\; \sum_{i=1}^{n} \beta_i a_i \geq \sum_{i=1}^{n} \beta_i b_i.$$

We next prove that $\succeq_{\mathbf{r}}$ satisfies Ordering, Continuity, Payoff Dominance, Baseline Independence, and Nepotism if and only if the focal individual is a quasi-inclusive fitness maximizer, as Theorem 2 asserts.

Proof of Theorem 2. It is straightforward to verify that if the focal individual is a quasi-inclusive fitness maximizer, then $\succeq_{\mathbf{r}}$ satisfies Ordering, Continuity, Payoff Dominance, Baseline Independence, and Nepotism.

Now, suppose that $\succeq_{\mathbf{r}}$ satisfies these five axioms. In view of Lemma 3, all that remains to be established is that, for all $j, k \in I$, the parameters are such that (i) $r_j > r_k$ implies $\beta_j > \beta_k$ and (ii) $r_j = r_k$ implies $\beta_j = \beta_k$.

Consider case (i) first. Suppose that there exist $j, k \in I$ such that $r_j > r_k$. Let $\mathbf{a}, \mathbf{b} \in F$ and $x > 0$ be such that $a_j = a_k =: a_0$, $b_j = a_0 + x$, $b_k = a_0 - x$, and $b_i = a_i$ for all $i \in I \setminus \{j,k\}$. Nepotism implies that $\mathbf{b} \succ_{\mathbf{r}} \mathbf{a}$. By Lemma 3 and the definition of $\mathbf{a}$, $\mathbf{b}$, and x,

$$\begin{array}{rcl} \mathbf{b} \succ_{\mathbf{r}} \mathbf{a} & \Leftrightarrow & \displaystyle\sum_{i=1}^{n} \beta_i b_i > \sum_{i=1}^{n} \beta_i a_i \\ & \Leftrightarrow & \beta_j b_j + \beta_k b_k > \beta_j a_j + \beta_k a_k \\ & \Leftrightarrow & (\beta_j + \beta_k) a_0 + (\beta_j - \beta_k) x > (\beta_j + \beta_k) a_0 \\ & \Leftrightarrow & (\beta_j - \beta_k) x > 0. \end{array}$$

Because $x > 0$, the last inequality implies that $\beta_j > \beta_k$.

The proof of case (ii) is similar. In this case, suppose that there exist $j, k \in I$ such that $r_j = r_k$. Defining $\mathbf{a}, \mathbf{b}$ as above, Nepotism implies $\mathbf{b} \sim_{\mathbf{r}} \mathbf{a}$. Replacing the inequalities with equalities in the displayed array, it follows that

$$\mathbf{b} \sim_{\mathbf{r}} \mathbf{a} \Leftrightarrow (\beta_j - \beta_k) x = 0.$$

Hence, $\beta_j = \beta_k$ because $x > 0$. □

Acknowledgements

We are grateful to Ken Binmore for comments and discussion. This research was supported by the European Research Council Seventh Framework Program (FP7/2007-2013), ERC Grant agreement no. 295449 (SO), the Gertrude Conaway Vanderbilt Chair in Economics (JAW), and the Fonds de Recherche sur la Société et la Culture du Québec and the Social Sciences and Humanities Research Council of Canada (WB).

References

Blackwell, D. A., Girshick, M. A., 1954. Theory of Games and Statistical Decisions. Wiley, New York.

Bossert, W., Qi, C. X., Weymark, J. A., 2013a. Extensive social choice and the measurement of group fitness in biological hierarchies. Biol. Phil. 28, 75–98.

Bossert, W., Qi, C. X., Weymark, J. A., 2013b. Measuring group fitness in a biological hierarchy: An axiomatic social choice approach. Econ. Phil., forthcoming.

Bossert, W., Weymark, J. A., 2004. Utility in social choice. In: Barberà, S., Hammond, P. J., Seidl, C. (Eds.), Handbook of Utility Theory. Volume 2: Extensions. Kluwer Academic Publishers, Boston, pp. 1099–1177.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection. John Murray, London.

d'Aspremont, C., 1985. Axioms for social welfare orderings. In: Hurwicz, L., Schmeidler, D., Sonnenschein, H. (Eds.), Social Goals and Social Organizations: Essays in Memory of Elisha S. Pazner. Cambridge University Press, Cambridge, pp. 19–76.

Dugatkin, L. A., 2007. Inclusive fitness theory from Darwin to Hamilton. Genetics 176, 1375–1380.

Fisher, R. A., 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Frank, S. A., 1998. Foundations of Social Evolution. Princeton University Press, Princeton, NJ.

Gardner, A., West, S. A., Wild, G., 2011. The genetical theory of kin selection. J. Evol. Biol. 24, 1020–1043.

Grafen, A., 2006. Optimization of inclusive fitness. J. Theor. Biol 238, 541–563.

Grafen, A., 2009. Formalizing Darwinism and inclusive fitness theory. Phil. Trans. R. Soc. Lond. B Biol. Sci. 364, 3135–3141.

Haldane, J. B. S., 1932. The Causes of Evolution. Longmans Green, London.

Haldane, J. B. S., 1955. Population genetics. In: Johnson, M. L., Abercrombie, M., Fogg, G. E. (Eds.), New Biology 18. Penguin, Harmondsworth, UK, pp. 34–51.

Hamilton, W. D., 1963. The evolution of altruistic behavior. Am. Nat. 97, 354–356.

Hamilton, W. D., 1964a. The genetical evolution of social behaviour. I. J. Theor. Biol. 7, 1–16.

Hamilton, W. D., 1964b. The genetical evolution of social behaviour. II. J. Theor. Biol. 7, 17–52.

McElreath, R., Boyd, R., 2007. Mathematical Models of Social Evolution: A Guide for the Perplexed. University of Chicago Press, Chicago.

Nowak, M. A., Tarnita, C. E., Wilson, E. O., 2010. The evolution of eusociality. Nature 466, 1057–1062.

Okasha, S., 2009. Individuals, groups, fitness and utility: Multi-level selection meets social choice theory. Biol. Philos. 24, 561–584.

Wagner, G. P., 2010. The measurement theory of fitness. Evolution 64, 1358–

1376.

Wenseleers, T., Gardner, A., Foster, K. R., 2010. Social evolution theory: A review of methods and approaches. In: Székely, T., Moore, A. J., Komdeur, J. (Eds.), Social Behaviour: Genes, Ecology and Evolution. Cambridge University Press, Cambridge, pp. 132–158.

Récents cahiers de recherche du CIREQ
Recent Working Papers of CIREQ

Si vous désirez obtenir des exemplaires des cahiers, vous pouvez les télécharger à partir de notre site Web http://www.cireqmontreal.com/cahiers-de-recherche

If you wish to obtain copies of the working papers, you can download them directly from our website, http://www.cireqmontreal.com/cahiers-de-recherche

08-2012 Bossert, W., Qi, C.X., J.A. Weymark, "Extensive Social Choice and the Measurement of Group Fitness in Biological Hierarchies", juillet 2012, 25 pages

09-2012 Bossert, W., K. Suzumura, "Multi-Profile Intertemporal Social Choice", juillet 2012, 20 pages

10-2012 Bakis, O., B. Kaymak, "On the Optimality of Progressive Income Redistribution", août 2012, 42 pages

11-2012 Poschke, M., "The Labor Market, the Decision to Become an Entrepreneur, and the Firm Size Distribution", août 2012, 29 pages

12-2012 Bossert, W., Y. Sprumont, "Strategy-proof Preference Aggregation", août 2012, 23 pages

13-2012 Poschke, W., "Who Becomes an Entrepreneur? Labor Market Prospects and Occupational Choice", septembre 2012, 49 pages

14-2012 Benchekroun, H., G. Gaudet, H. Lohoues, "Some Effects of Asymmetries in a Common Pool Natural Resource Oligopoly", août 2012, 24 pages

15-2012 Ehlers, L., B. Klaus, "Strategy-Proofness Makes the Difference : Deferred-Acceptance with Responsive Priorities", septembre 2012, 32 pages

16-2012 Bossert, W., C.X. Qi, J.A. Weymark, "An Axiomatic Characterization of the MVSHN Group Fitness Ordering", septembre 2012, 20 pages

17-2012 Ruge-Murcia, F., "Skewness Risk and Bond Prices", mai 2012, 41 pages

18-2012 Amarante, M., M. Ghossoub, E. Phelps, "Contracting for Innovation under Knightian Uncertainty", septembre 2012, 37 pages

19-2012 Bossert, W., C. D'Ambrosio, "Proximity-Sensitive Individual Deprivation Measures", décembre 2012, 15 pages

01-2013 Bossert, W., Y. Sprumont, "Every Choice Function is Backwards-Induction Rationalizable", janvier 2013, 15 pages

02-2013 Amarante, M., "Conditional Expected Utility", février 2013, 19 pages

03-2013 Benchekroun, H., F. Taherkhani, "Adaptation and the Allocation of Pollution Reduction Costs", mai 2013, 33 pages

04-2013 Bossert, W., H. Peters, "Single-Basined Choice", juin 2013, 15 pages