Journal of Theoretical Biology 350 (2014) 24–31

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

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, Canada H3C 3J7

HIGHLIGHTS

- We characterize the principle of inclusive fitness maximization axiomatically.
- We find behavioural conditions that are necessary and sufficient for an individual to maximize its inclusive fitness.
- We show formal links between inclusive fitness theory and rational choice theory.

ARTICLE INFO

Article history:
Received 24 May 2013
Received in revised form
2 February 2014
Accepted 3 February 2014
Available online 11 February 2014

Keywords:
Hamilton's rule
Inclusive fitness
Kin selection
Rational choice

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) for the case in which phenotypic effects are additive. 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.

1. Introduction

A central tenet of inclusive fitness theory is that a trait may be selected for even if it involves some sacrifice to an individual's personal fitness, provided that it sufficiently enhances the reproductive success of genetically related individuals. Genetic relatedness between social partners can arise for various reasons, in particular kinship. 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), or Wenseleers et al. (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 later expressed in print (see Haldane, 1955, p. 44). However, it was 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

* Corresponding author. Tel.: +44 117 928 7829.
E-mail addresses: Samir.Okasha@bristol.ac.uk (S. Okasha), john.weymark@vanderbilt.edu (J.A. Weymark), walter.bossert@umontreal.ca (W. Bossert).

http://dx.doi.org/10.1016/j.jtbi.2014.02.002
0022-5193

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 et al., 2010). (Birch (forthcoming) provides an illuminating discussion of the arguments for and against the claims made by Nowak et al. (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 et al. (2011), among others.

In this paper, 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 for the case in which phenotypic effects are additive. Consider a focal individual and the set of other individuals who might be affected by this individual's actions. At a given point in time, each of the latter individuals stands in a given 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 as follows: 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 et al. (2013a,b) to evaluate alternative measures of group fitness in hierarchically structured populations. This paper 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.

Our model is not evolutionary; rather it is behavioural. Our aim is to characterize mathematically a certain pattern of behaviour that organisms might exhibit, namely inclusive fitness maximization, in terms of the properties of a binary preference relation. We do not assert that the evolutionary process will necessarily lead organisms to exhibit the behaviour in question, or that it will 'tend' to do so, or that the behaviour, if it evolves, will be stable against mutation; and we do not study the conditions under which an allele coding for the behaviour will be favoured by natural selection. To address these questions would require constructing an explicit evolutionary model and studying its evolutionary dynamics. There is a large literature addressing these questions, and we are not attempting to contribute to it. Rather, our aim is different, namely to supply an alternative mathematical characterization of inclusive fitness maximizing behaviour.

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 discuss the significance of our results 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 $r_i \in \mathbb{R}$ denote the relatedness of the focal individual to individual i, with higher values denoting a closer degree of relatedness, where $r_1 = 1$. Thus, the set I has an associated relatedness profile $\mathbf{r}=(r_1,\ldots,r_n) \in 1 \times \mathbb{R}^{n-1}$. At a particular point in time, the profile $\mathbf{r}$ is taken as given. (However, we make no assumption about what determines $\mathbf{r}$; it may have ecological as well as genealogical determinants.) If relatedness depends on the evolving trait, then at a subsequent point in time the relatedness profile $\mathbf{r}$ will be different; and our analysis will apply again at that later time *modulo* the new relatedness profile.

In Hamilton's original papers, relatedness was defined as the probability that actor and recipient share genes that are 'identical by descent', which is determined by their genealogical relationship; this implies that $r_i \in [0,1]$. However later work, by Hamilton and many others, has shown that the relatedness that matters to inclusive fitness theory is a more abstract measure of genetic similiarity (typically, the regression of recipient genotype on actor genotype); for discussion, see Michod and Hamilton (1980), Frank (1998), Grafen (2006), or Rousset (2004, chapter 7). This means that, in principle, the relatedness co-efficient can assume any real value, including negative values, and that is why we assume $r_i \in \mathbb{R}$ rather than $r_i \in [0,1]$ for $i \neq 1$. Our formalism is deliberately neutral with respect to the precise definition of relatedness employed, which may be different in different evolutionary models.

At a given point in time, 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