
Behavioural structure and the evolution of cooperation

R. E. MICHOD AND M. J. SANDERSON

Introduction

Most kin selection and group selection models of social behaviour assume that differences in an individual's behaviour result only from differences in an individual's genotype. One of the main conclusions obtained from such models is that population structure among interacting genotypes is necessary for social behaviours, which accrue some cost to the individuals performing them, to spread. However, this necessity of genetic structure stems from the simplifying assumptions concerning the relationship between genotype and behaviour made in these models. In a situation in which individuals have behavioural flexibility and choose from a set of behaviours according to some rule (which may itself be genetically inherited), it is no longer necessary that there be genetic structure for costly behaviours to evolve. Indeed, genotypes can interact completely at random, yet altruistic or cooperative strategies may still be favoured (Brown, Sanderson & Michod, 1982). In order for this to occur, there must be some structure in the associations between the *behaviours* themselves (see appendix). In general, the requirement of genetic structure is replaced by a requirement of *behavioural structure*, so that the behaviours are encountered in frequencies different from their overall frequency in the population. In this chapter, we use a measure of behavioural structure (defined below) to relate the main theories for the evolution of social behaviour.

Discussion of terms

For lack of a better term, we use the term 'behaviour' to refer to the actions and activities which mediate a fitness interaction between two individuals. Consequently, the role of behavioural structure discussed below applies to fitness interactions in general, no matter how they are mediated. In particular, the ideas developed here are not restricted to animals with

overt behaviours, but may also apply, for example, to populations of interacting plants.

The term 'altruism' refers to a behaviour which increases the genotypic fitness of other individuals, who are termed the recipients, while decreasing the genotypic fitness of the individual exhibiting the behaviour, who is termed the donor. Consequently, the 'cooperate' behaviour of the 'Prisoner's Dilemma' game (Axelrod & Hamilton, 1981; see Table 1 below) is an altruistic behaviour. In this chapter we study the evolution of a population composed of individuals which display two behaviours, termed 'cooperate' and 'defect', and denoted c and d respectively. Let C be the additive effect of a cooperative behaviour on the individual fitness of the donor and B the additive effect on the individual fitness of the recipient. Defection is assumed to have no effect on the fitness of either interactant.

The conditional distribution of interactions among *genotypes* is the basis of most models of kin selection, and allows different kin selection models to be related to each other (Abugov & Michod, 1981; Michod, 1982). We propose here that an analogous measure defined at the level of the behavioural interactions themselves can be used to relate the various evolutionary mechanisms for the evolution of cooperation, such as kin selection, assortment, reciprocation, and group selection. Let u_{ij} be the probability that an individual exhibiting behaviour j interacts with an individual exhibiting behaviour i . Let u_i and u_j be the frequencies of behaviour i and j in the population. There is structure in the distribution of behaviours to the extent that u_{ij} differs from u_i . Consequently, u_{ij} measures the behavioural structure in the population.

General condition

In this chapter we use the measure of behavioural structure defined above, $u_{c|c}$, to relate the main theories for the evolution of cooperation. We do this by expressing in the following general form the various conditions for increase of a cooperative trait which have been derived by these theories:

$$C/B < k u_{c|c}. \quad (1)$$

To express the various conditions for increase of a cooperative trait in the form of condition (1), k is defined operationally by first solving for $u_{c|c}$ in the various models and then letting k be whatever is necessary for condition (1) to hold. In all cases considered, k thus operationally defined can be interpreted as the fraction of $u_{c|c}$ which is due to the proximate mechanisms under study, such as family structure, assortment, reciprocation or group structure. Consequently, we refer to this component of behavioural structure as being 'directed' by the various mechanisms under study. In each case, k depends upon a set of constants specific to the mechanism

under study and on the mean frequency of cooperators in the population. Although the particular form of k varies from theory to theory, k always equals the magnitude of the directed component of behavioural structure. By representing the various theories in this common framework, the similarities and differences between them can be clearly represented and understood.

Kin selection

It is well known (see Michod, 1982, for review) that the condition for increase of an altruistic trait in a family-structured population is given by Hamilton's rule (1964),

$$C/B < 1/2. \quad (2)$$

This condition applies globally if selection is weak, there is no over- or underdominance and the fitness effects are additive (Michod & Abugov, 1980). If selection is strong, condition (2) gives the condition for increase of an altruistic trait when rare or when common for both the additive and multiplicative (not considered here) models of fitness (Cavalli-Sforza & Feldman, 1978).

We now relate condition (2) to behavioural structure and condition (1). To help fix ideas, consider a simple model of a haploid, sexual population. Assume a single locus with two alleles A and a which cause a difference in behaviour so that A individuals cooperate and a individuals defect on every interaction no matter how their partner behaves. There are three families possible in a sexual, haploid population designated by the genotypes of the parents: $A \times A$, $A \times a$, and $a \times a$. The conditional distribution of interactions between the cooperating behaviours can be calculated as follows. Given that an individual is c , then one of its parents must have been c . The other parent would have been c with probability u_c and d with probability $1 - u_c$, where u_c is the frequency of A in the population. The probability that an offspring receives its allele from any particular parent is, of course, $1/2$. Consequently, the probability that a sib of a c individual is also c is

$$u_{c|c} = 1/2 + u_c/2. \quad (3)$$

Hamilton's condition (2) can now be rewritten in the form of condition (1) with

$$k = \frac{1}{1 + u_c}. \quad (4)$$

In Eqn (4), k is the fraction of $u_{c|c}$ which is specifically due to identity by descent and the family structure. Note that the conditional distribution of interactions between cooperators (Eqn 3) contains two components. The first component on the right-hand side of Eqn (3), $1/2$, stems directly from

identity by descent. This component contributes $1/(1+u_c)$ to the total behavioural structure. The second component in Eqn (3), $u_c/2$, stems from the population frequency of the cooperators and contributes $u_c/(1+u_c)$ to the total behavioural structure. In condition (1) using Eqns (3) and (4), the total behavioural structure is multiplied by the fraction of it which is directed to give the well-known $1/2$ in Hamilton's rule (Eqn 2). Our reason for representing Hamilton's rule in the more complex form of condition (1) is to relate kin selection to the other main evolutionary paths to cooperation discussed below.

Assortment of encounters

Eshel & Cavalli-Sforza (1982) have studied a model for the evolution of cooperation which assumes that individuals are either always c or always d and that individuals interact with other like types with probability m and at random with probability $1-m$. They discuss briefly several proximate mechanisms by which individuals may assort with other like types. When interpreted in terms of additive cost and benefit, their condition (3.2) for increase of cooperation becomes

$$C/B < m. \quad (5)$$

When interpreted in terms of the conditional distribution of interactions between cooperators, condition (5) becomes condition (1) with

$$k = \frac{m}{m + (1-m)u_c}. \quad (6)$$

In this case, a c individual expects $m + (1-m)u_c$ encounters with other c individuals. However, of these interactions, only m are due to assortment. Thus $m/[m + (1-m)u_c]$ in Eqn (6) represents the fraction of a c 's encounters which are with individuals who cooperate because of assortment.

Reciprocation

The theory of reciprocation (Trivers, 1971), as extended by Axelrod & Hamilton (1981), assumes that individuals can modify their behaviour through information they acquire about their environment. A simple learning rule which incorporates the basic idea of reciprocation is 'Tit for Tat' (or TFT), according to which if an individual recognizes its present partner from a previous interaction, then the individual behaves now as did its partner during that previous encounter. However, if the individual does not recognize its partner, then the individual cooperates. Axelrod & Hamilton (1981) and Brown, Sanderson & Michod (1982) studied the evolution of the TFT reciprocating strategy in competition with a strategy of total defection (called ALL D). Define r as the frequency of TFT

individuals in the population, with $1-r$ the frequency of ALL D individuals. Let v be the probability that an individual is recognized from a previous encounter. For the additive model considered here, Brown, Sanderson & Michod (1982, Eqn (30), in which $\beta = \alpha(1-v)$) derived the following condition for increase of the reciprocating strategy

$$C/B < vu_{c|c}, \quad (7)$$

which is in the form of condition (1). Consequently, in this case

$$k = v, \quad (8)$$

and so the fraction of behavioural structure which is directed is simply the probability of recognizing an individual. As shown in Brown *et al.* (1982), condition (7) applies to several multi-partner models of interaction and not just the single-partner model considered by Axelrod & Hamilton (1981). It has also been shown (Brown *et al.*, 1982) that condition (7) applies to a single locus, two allele, diploid population genetic model in which the one homozygote has phenotype TFT, the other homozygote has phenotype ALL D, and the heterozygote behaves according to the TFT rule with probability h and ALL D with probability $1-h$.

For a population composed of TFT and ALL D individuals, $u_{c|c} = r^2/u_c$. This may be derived using the definition of conditional probability. By definition of TFT and ALL D, a $c-c$ interaction occurs if and only if a TFT-TFT interaction occurs. Thus, the joint probability of $c-c$ interactions is r^2 in a population in which individuals or genotypes interact at random. There are again two components to this behavioural structure. Because of the nature of the TFT learning rule, all encounters between two TFT individuals result in mutual cooperation. However, only a fraction v of these are directed in the sense that they are due to *recognition* and, hence, reciprocation *per se*. The remaining fraction, $1-v$, of $u_{c|c}$ arises simply because all TFT individuals cooperate indiscriminately on their first encounter with a stranger.

Group selection

We now apply these ideas of behavioural structure to D. S. Wilson's (1980) 'structured deme' model of group selection, which studies the evolution of two types such as c and d as defined above. Wilson (1980, Chapter 2) has also discussed the relations between behavioural structure and kin selection. Individuals are assumed to interact in the context of local subpopulations of constant size, termed 'trait groups', and to reproduce by a process, such as asexual haploid reproduction, in which the Darwinian fitness of a type is directly related to its frequency in the next generation. Let σ^2 be the variance in the frequency of c types between trait groups and u_c the frequency of c types in the total population. Let $u_{c,i}$ denote the frequency of c

in the i th local population. The fitnesses of the two types in a local population are then

$$\begin{aligned} W_{c,i} &= 1 - C + u_{c,i}B \\ W_{d,i} &= 1 + u_{c,i}B. \end{aligned} \quad (9)$$

For Wilson's model, $u_{c|c}$ and $u_{c|d}$ (identical to Wilson's 'experienced frequencies') are given by the following equations (in which $E[x]$ denotes the expected value of x):

$$u_{c|c} = \frac{E[u_{c,i}^2]}{u_c} = u_c + \frac{\sigma^2}{u_c} \quad (10a)$$

$$u_{c|d} = \frac{E[u_{c,i}u_{d,i}]}{u_d} = u_c - \frac{\sigma^2}{u_d} \quad (10b)$$

Following Wright (1951), define F_{ST} as the correlation between types picked at random in the local subpopulations relative to types picked at random from the total population,

$$F_{ST} = \frac{\sigma^2}{u_c u_d}. \quad (11)$$

The correlation F_{ST} (11) is the ratio of the actual variance in frequency of c types to the maximal possible variance if all local subpopulations were fixed for one type or the other. Using Eqns (9)–(11), it can be shown that the expected fitness of the c type will be greater than the expected fitness of the d type if

$$\frac{C}{B} < F_{ST}. \quad (12)$$

To show the relationship of condition (12) to condition (1), write Eqn (10) as (using Eqn(11)):

$$u_{c|c} = F_{ST} + (1 - F_{ST})u_c \quad (13a)$$

$$u_{c|d} = (1 - F_{ST})u_c. \quad (13b)$$

There are two components of these conditional probabilities which represent the behavioural structure. The second component on the right-hand side of Eqn (13a) is experienced by both types and so cannot give rise to any difference in fitness between the types. However, the first component on the right-hand side of Eqn (13a), F_{ST} , is directed to the c type alone. Consequently,

$$k = \frac{F_{ST}}{F_{ST} + (1 - F_{ST})u_c}. \quad (14)$$

Using the definitions given in Eqns (13a) and (14), it is easy to see that $ku_{c|c} = F_{ST}$, which shows that condition (12) is again a special case of condition (1).

Discussion

Through simple algebraic rearrangements, we have expressed in the common form of condition (1) the conditions for evolution of cooperation derived in the theories of kin selection, assortment, reciprocity, and Wilson's structured deme model of group selection. Condition (1) predicts that cooperative behaviours which accrue some cost in individual fitness will increase in frequency if their cost/benefit ratio is less than the directed component of behavioural structure, $ku_{c|c}$. Behavioural structure measures the degree to which a cooperative behaviour is associated with other cooperative behaviours. This concept shifts attention from the level of the individual or genotype to the level of the behaviours themselves. The fraction of the total behavioural structure which is directed is k , which depends on the specific mechanisms by which behavioural structure is produced in the four theories studied (Eqns 4, 6, 8 and 13). The non-directed component of the structure does not enter into the condition for selection, because this component is the same for all types in the population and so does not produce any differential fitness effect.

The fraction of the behavioural structure which is directed, k , has a different character in the four theories. In intersibling kin selection, k (Eqn 3) has a minimum value of 1/2, while under the TFT rule of reciprocity, assortment, or group selection, it can become arbitrarily small, as the probability of recognizing an individual, v , the probability of assortative meeting, m , or the variance between local subpopulations, σ^2 , get small.

In models which assume that all differences in behaviour are caused by differences in genotype, behavioural structure can only be generated by genetic population structure. This is the case with most models of kin and group selection as well as with Eshel & Cavalli-Sforza's (1982) model of assortment discussed above. However, in the models of reciprocity discussed above, behavioural structure is generated by the learning process. In this case, genotypes may interact completely at random, yet costly social behaviours may still evolve. In such models, differences in genotypic fitness are generated by differences in the learning rules by which genotypes modify their behaviour based on previous experience.

As mentioned in the introduction, the conclusion that genetic structure is necessary for costly, social traits to evolve is a direct consequence of the assumption of a one-to-one correspondence between behaviour and genotype usually made in sociobiological models. Consequently, genetic structure need not be a feature of sociality in the broad sense. In fact, in organisms with some degree of behavioural plasticity, it is more likely that behaviours arise due to an interaction between heritable traits and environmental events (specifically including interactions with conspecifics and learning). This chapter has attempted to show that in such situations

Table 1. Payoff matrix for the *c* and *d* behaviours with additive costs and benefits

Payoff to behaviour below when encountering behaviour to right	<i>c</i>	<i>d</i>
	<i>c</i>	$1+B-C^a$
<i>d</i>	$1+B$	1

^a The parameters *B* and *C* are the additive effects on the fitnesses of recipient and donor, respectively. See text for further explanation.

the appropriate focus for research is the structure among behaviours in a population and not the genetic structure. Indeed, behavioural structure is necessary for the evolution of costly, social behaviours in all the models studied here. This in no way implies that evolution of these behaviours is somehow removed from genetic evolution, but it does mean that the observable manifestations of such genetic evolution may emerge on the behavioural level independent of considerations such as genetic population structure.

We thank Bob Abugov, Renee D. Rusler, Paul Sherman and Marcy Uyenoyama for discussion and comment on earlier versions of this manuscript. This work was supported, in part, by NSF grant DEB-8118248 to R.E.M.

Appendix

In this appendix we consider the effects of random interactions on individual fitness. We do not address the process by which the traits which mediate the interactions are transmitted from one generation to another. The transmission process is, of course, an integral part of the evolution of any trait, behaviour included. However, our purpose here is more limited in scope than to develop a complete evolutionary description of any particular behaviour. We only wish to illustrate the importance of behavioural structure to individual fitness, realizing that the transmission process will further affect the evolution of the trait in ways not considered here (see, for example, Michod, 1984).

As above, let u_{ij} be the probability that an individual exhibiting behaviour *j* interacts with another individual exhibiting *i*. Let u_i be the frequency of behaviour *i* in the total population. Consider the case assumed above of two behaviours *c* and *d* with additive costs and benefits to fitness. The interactions between individuals result in one of four possible fitness payoffs given in Table 1, which satisfy the conditions of the Prisoner's

Dilemma game (see, for example, Axelrod & Hamilton, 1981; or Brown *et al.*, 1982).

The expected payoffs to individuals exhibiting behaviours *c* and *d* are

$$\begin{aligned} W_c &= u_{c|c}(1+B-C) + (1-u_{d|c})(1-C) \\ W_d &= u_{c|d}(1+B) + (1-u_{d|d}) \times 1. \end{aligned} \quad (A1)$$

If encounters among behaviours occur at random, then

$$\begin{aligned} u_{c|c} &= u_c \\ u_{d|c} &= 1 - u_c \\ u_{c|d} &= u_c \\ u_{d|d} &= 1 - u_c. \end{aligned} \quad (A2)$$

Using Eqns (A2),

$$\begin{aligned} W_c &= u_c(1+B-C) + (1-u_c)(1-C) \\ W_d &= u_c(1+B) + (1-u_c). \end{aligned} \quad (A3)$$

By inspection, $W_d > W_c$ in all cases. During any interaction an individual adopting *c* can expect to receive a lower payoff than one adopting *d*. In many cases of genetic transmission, this should be sufficient to prevent altruism from spreading in a population. However, in some cases of cultural transmission (Cavalli-Sforza & Feldman, 1981), altruism could still be favoured. In such cases, the success of altruism would be due to a bias in the transmission process due to cultural reasons.

References

- Abugov, R. & Michod, R. (1981). On the relation of family structured models and inclusive fitness models for kin selection. *Journal of Theoretical Biology*, **88**, 733-42.
- Axelrod, R. & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, **211**, 1390-6.
- Brown, J., Sanderson, M. & Michod, R. (1982). Evolution of social behavior by reciprocation. *Journal of Theoretical Biology*, **99**, 319-39.
- Cavalli-Sforza, L. L. & Feldman, M. (1978). Darwinian selection and 'altruism'. *Journal of Theoretical Population Biology*, **14**, 263-81.
- Cavalli-Sforza, L. L. & Feldman, M. (1981). *Cultural Transmission and Evolution*. Princeton: Princeton University Press.
- Eshel, I. & Cavalli-Sforza, L. L. (1982). Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences, USA*, **79**, 1331-5.
- Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology*, **7**, 1-52.
- Michod, R. E. (1982). The theory of kin selection. *Annual Review of Ecology and Systematics*, **13**, 23-55.
- Michod, R. E. (1984). Genetic constraints on adaptation with special reference to social behavior. In *The New Ecology: Novel Approaches to Interactive Systems*,

- ed. P. W. Price, C. N. Slobodchikoff & W. S. Gaud, pp. 253–78. New York: Wiley.
- Michod, R. E. & Abugov, R. (1980). Adaptive topography in family-structured models of kin selection. *Science*, **210**, 667–9.
- Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, **46**, 35–57.
- Wilson, D. S. (1980). *The Natural Selection of Populations and Communities*. Menlo Park, California: Benjamin/Cummings.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, **15**, 323–54.

Reprinted from *Evolution*—essays in honour of John Maynard Smith
Edited by J. J. Greenwood and M. Slatkin
© Cambridge University Press 1985
Printed in Great Britain at the University Press, Cambridge