Hamilton's Rule in Non-Additive Games

Simon J. Tudge*[1], Markus Brede[1], Richard A. Watson[1]
and Miguel Gonzalez[1]
[1]The University of Southampton, SO17 1BJ, UK
*Corresponding Author
Email: sjt4g11@soton.ac.uk

May 13, 2014

Abstract

Recently a number of authors have questioned both the validity and utility of inclusive fitness. One particular claim is that Hamilton's rule applies only to additive games. Additive games represent a vanishingly small subset of all games and do not capture a number of interesting qualitative behaviours which are present in non-additive games. Thus, if these criticisms were correct, inclusive fitness would be a severely limited theoretical tool. We show these criticisms are not valid by demonstrating that any symmetric game can be transformed into an additive payoff matrix in such a way that the action of selection remains unchanged. The result comes with a caveat, however, which is that terms in the payoff matrix must themselves be frequency dependent. Despite this, we demonstrate the utility of inclusive fitness by means of applying Hamilton's rule to two such non-additive games. The central claim of inclusive fitness is that relatedness is the key to cooperation, we show that this remains true even for non-additive games.

Keywords

Inclusive fitness, evolutionary game theory, non-additive games, evolution of cooperation, synergy.

Highlights

- We review the notion of an additive game in the context of evolutionary game theory.
- We show how inclusive fitness can be applied to additive games.
- We show how any non-additive evolutionary game can be transformed into an additive one without altering the dynamics of selection.

- We apply the notion of inclusive fitness to two general classes of non-additive games.

1 Social Evolution and Inclusive Fitness

A social trait is any trait which has a fitness altering effect on other members of the population in question. Typical examples might include: fighting behaviour, sexual strategies or signaling (see for example [10, 22]). In the case of social evolution the fitness of a particular trait is not an absolute (or a function of a static environment) but is dependent on the frequencies with which other types of individuals are present in the population. Thus, it seems that the problem of social evolution is a fundamentally more difficult task for theorists than the more basic problem of frequency-independent selection.

The problem of the evolution of cooperation has been a particularly prominent research question within the field of social evolution. The research program attempts to explain the seemingly paradoxical observation that some organisms forgo reproductive potential in order to increase the fitness of others. Prominent examples of this are sterile castes in the eusocial insects [8], stalk cells in slime moulds [19] and somatic cells in multicellular organisms [11]. Many very plausible explanations and well developed theories exist which can explain cooperation in the biological world; the most prominent of which is inclusive fitness (I.F.) theory [7, 6]. This paper is primarily concerned with IF and how this interpretation can be arrived at starting from the assumptions typically made in evolutionary game theory.

I.F. partitions fitness into two terms, the first being cost and the second being benefit weighted by the relatedness between the donor and the recipient. The inclusive fitness of an individual is $rb - c$ (see Grafen [5] for potential pitfalls in this approach). The meaning of each symbol is summarised as follows:

1. Relatedness, r, which measures the extent to which interactions are correlated between social strategies.

2. Cost, c, which measures the extent to which that certain behaviour decreases the actor's expected number of offspring.

3. Benefit, b, which measures the extent to which the expected number of offspring of the recipient of the behaviour increases.

One major cause of controversy of late has been over the nature of the cost and benefit terms. Some authors, most prominently Nowak et. al. [15], have claimed that such a decomposition of fitness is hardly ever valid. This is because in many instances there are no appropriate quantities in their models to equate with Hamilton's c and b. Particularly, if interactions are synergistic, costs and benefits by their very nature cannot be ascribed to any one action but are the combined outcome of two (or more) actions being performed together [16]. Games which are additive describe straightforward interactions in the absence of synergy, such that each strategy or behaviour can be construed of as a behaviour which incurs a constant cost and donates a constant benefit to any other individual with which it interacts.

In other words the incurred cost and donated benefit are independent of the phenotype of the recipient of the action. Clearly this is a very specific assumption to make and there is no *a prior* reason to think that many real life systems would have this property. In such a situation the application of IF is straightforward and uncontroversial, however, such games represent a vanishingly small proportion of all possible games.

The central result of this paper is that any payoff matrix (additive or otherwise) can be transformed into a payoff matrix which *is* additive, in such a way that the action of selection remains unchanged. Due to this transformation the I.F. approach remains valid even for non-additive games. The bottom line results are two simple formulae for the appropriate costs and benefits to use in an I.F. decomposition. These formulae are presented in terms of simple matrix operations of the original payoff matrix and are valid for an arbitrary number of strategies.

We will briefly review the key assumptions of evolutionary game theory. We will then show how this formalism can be neatly extended to include population structure. We then review the meaning of an additive game and show how a notion of inclusive fitness can very easily be arrived at in such a case. The main result of the paper then follows, which shows how any non-additive game can be transformed into an additive one so that the inclusive fitness decomposition remains valid. We apply this formalism to both the stag hunt and snowdrift game, which are both non-additive games, in order to demonstrate the utility of inclusive fitness. The central claim of inclusive fitness is that relatedness facilitates the evolution of cooperation, we show that this remains true in general.

2 Payoff and Assortment

Evolutionary game theory is often coupled with equations of motion such as the replicator equation [20] to give a dynamic account of selection. However, for our purposes we do not need a full dynamic account of selection, we are simply interested in whether or not a given strategy will increase or decrease in frequency. By assumption a strategy will increase in frequency if it has a higher than average fitness and it is present in non-zero frequencies in the population. A strategy i has a payoff π_i and a relative frequency x_i (such that frequencies sum to one). The average payoff is denoted by $\bar{\pi}$. In a well-mixed population individuals play a game with a random member of the population, in which case payoff can be calculated via:

$$\pi_i = \sum_j x_j M_{ij} \tag{1}$$

Where M_{ij} is the payoff i receives upon encountering a j.

In all cases average payoff is calculated via $\bar{\pi} = \sum_i x_i \pi_i$.

It has often been remarked upon that structuring of interactions is the key to the evolution of cooperation. Specifically positive assortment facilitates the evolution of cooperation and altruism [1, 18, 2, 4, 12]. One particular way of seeing this is to realise that in a positively assorted population the benefits of cooperation fall

disproportionately upon those who cooperate and thus it may become rational to do so. Equation (1) does not take into account any structuring of interactions and therefore lacks some generality. More recently attempts have been made to incorporate population structure into the framework of evolutionary game theory (in particular Van Veelen [21] whose approach we build on here).

More generally than equation (1) the payoff to an individual i is given by the probability that it meets an individual j, multiplied by the payoff received against a j summed over all possible js. That is:

$$\pi_i = \sum_j P_{ij} M_{ij} \tag{2}$$

where P_{ij} should be read as the probability that an individual meets a j given that it is of type i. The calculation of average payoff is unaltered. The well-mixed condition is a special case of the above in which $P_{ij} = x_j$ and is therefore independent of i. Of particular interest to social evolution theory is positive assortment; whereby strategies meet their own types more often than would be expected from random interactions.

A simple model captures the key features of assortment. Consider a focal individual of type i, with probability α it is paired with a clonally related individual and with probability $1 - \alpha$ it is paired with an individual chosen at random from the population. Under this formulation of assortment the parameter α is equivalent to Hamilton's r (see [3] for a proof). The matrix P can thus be written as:

$$P_{ij} = \begin{cases} (1-\alpha)x_j + \alpha & j = i \\ (1-\alpha)x_j & j \neq i \end{cases} \tag{3}$$

This can be expressed more concisely using the delta matrix: δ for which $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. Thus we may write:

$$P_{ij} = (1-\alpha)x_j + \alpha\delta_{ij} \tag{4}$$

2.1 Additivity

One of the main charges Nowak et al. bring against Hamilton's rule is that it only applies to additive games; we show here that this is not true. Additive games have a feature known as *equal gains from switching* [13]. Given a particular pairwise interaction if we were to hypothetically change the strategy of a focal player then the resulting change in payoff for that player would be independent of the strategy of its partner.

Formally the payoff matrix must satisfy:

$$M_{ik} - M_{il} = M_{jk} - M_{jl} \quad \forall i, j, k, l \tag{5}$$

A general two-player game is given by the payoff matrix:

$$M = \begin{pmatrix} R & S \\ T & P \end{pmatrix} \tag{6}$$

For this payoff matrix to be additive it is necessary that: $R - S = T - P$. One can see that this is a rather special condition and is not likely to be met for a random payoff matrix.

Any additive game can be represented as a donation game in which an individual may pay a certain cost to bestow a benefit upon another individual (Crucially the benefit may outweigh the cost). If strategy i costs an individual c_i units of fitness to perform, and donates b_i units of fitness to its partner then such a game can be written as:

$$M_{ij} = b_j - c_i \tag{7}$$

this satisfies equation (5) and is therefore additive. Furthermore, it will always be possible to arbitrarily choose one strategy to have zero cost and benefit. This is because payoffs are relative, and the direction of selection is unaltered upon addition of a constant to the payoff matrix.. By subtracting $b_1 - c_1$ from M we are left with a payoff matrix in which the top left corner is zero (or any other diagonal element of our choosing). Then all other cost and benefit terms can be considered as being relative to strategy one.

In the case of an assorted population it is possible to arrive at the notion of inclusive fitness and hence to Hamilton's rule.

Payoff to individual i is:

$$\begin{aligned} \pi_i &= \sum_j P_{ij} M_{ij} && (8) \\ &= \sum_j \{(1-\alpha)x_j + \alpha\delta_{ij}\}(b_j - c_i) && (9) \\ &= (1-\alpha)\,\underline{b}.\underline{x} + \alpha b_i - c_i && (10) \end{aligned}$$

As payoff is an inherently relative concept it may be defined up to an arbitrary constant. We therefore subtract the constant term and arrive at: $\pi_i = \alpha b_i - c$ which is exactly the inclusive fitness of individual i. Hamilton's rule, in its simplified form, simply asks when cooperators increases in frequency with respect to defectors, where a cooperator donates a fitness benefit b at a cost of c to itself and the defective strategy donates no benefit at no cost. Thus, the payoff of a cooperator is $\pi_c = \alpha b - c$ and the defector $\pi_d = 0$ and thus cooperation increases in frequency if: $\alpha b - c > 0$, which is Hamilton's rule if one equates α with r.

2.2 ST Space

Santos et. al. [17] introduced a powerful tool for considering two-by-two games by showing that we may set $R = 1$ and $P = 0$ without loss of generality. Thus the space of all possible two-player games is two-dimensional (named ST space). They were able to show how the space is composed of 4 games with qualitatively different behaviour (prisoner's dilemma, harmony game, snowdrift game and stag hunt game). In ST space additive games satisfy the constraint T = 1-S and thus lie on a line through the space of all two-player games which passes through the prisoner's dilemma and the harmony game only (see figure 1). The prisoner's dilemma

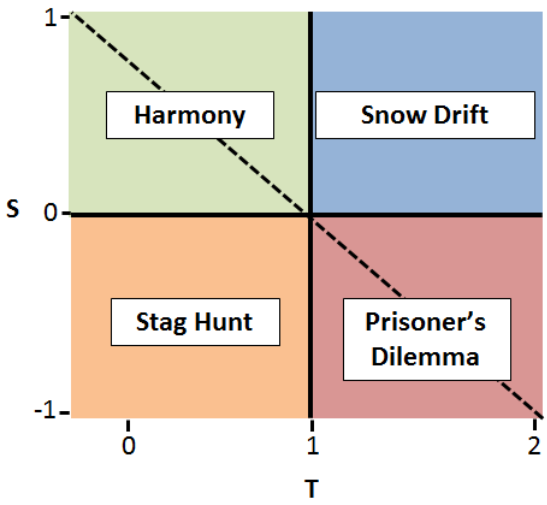


Figure 1: ST space: the space of all possible two player games. The four quadrants correspond to 4 qualitatively different types of dynamics. The principle diagonal, represented with the dashed line, is the one-dimensional subset of games which are additive, and can hence be described in terms of constant costs and benefits.

and harmony game both have pure equilibria. The snowdrift game has a mixed equilibrium and the stag hunt game is bistable. These two more interesting types of behaviour are features of the non-additive nature of the payoff matrix.

3 Non-Additive Games

The central result of this paper is that any possible payoff matrix can be transformed into an additive matrix in such a way that the direction of selection is unaltered. This is in agreement with the inclusive fitness research programme, which claims that Hamilton's rule is general. The caveat in this approach is that the cost and benefit terms are now frequency dependent, that is they can in general depend upon the state of the population $\underline{x}$. What follows is a simple procedure for determining the appropriate costs and benefits:

$$\begin{aligned} \pi_i &= \sum_j \left[(1-\alpha)\, x_j + \delta_{ij}\alpha\right] M_{ij} && (11)\\ &= (1-\alpha)\,(M.x)_i + \alpha M_{ii} && (12)\\ &= (M.x)_i + \alpha(M_{ii} - (M.x)_i) && (13)\\ &= \alpha B_i(x) - C_i(x) && (14) \end{aligned}$$

with $C_i(x) = -(M.x)_i$ and $B_i(x) = M_{ii} - (M.x)_i$. The cost and benefit terms are chosen in such a way so that they fit the appropriate form for inclusive fitness.

For some intuition it is instructive to look again at the general form of the two player game in equation (6). We define all costs and benefits relative to that of the second player, so that: $C(x) = c_1 - c_2$. $M.x = (Rx + S(1-x), Tx + P(1-x))$ and hence $C(x) = x(T-R) + (1-x)(P-S)$. This term is the expected gains from switching in the well-mixed case. That is, given that ones partner is chosen at random what is the average change in payoff upon changing from strategy 2 to strategy 1. In a similar manner we define $B(x) = b_1 - b_2$. Then: $B(x) = R - P + x(T-R) + (1-x)(P-S)$, the intuition behind this term is less straightforward. $R-P$ is the difference between strategy one and strategy two both playing against themselves. The remainder of the term is again the expected gains from switching in the well-mixed case. The benefit term therefore represents the gains from self interactions minus the gains from switching assuming random interactions.

More generally the cost in Hamilton's rule is simply the expected payoff in the well-mixed case, the benefit is the difference between the self payoff and the expected well-mixed payoff. Relatedness simply measure the extent to which one is likely to meet a like type above that which would be expected from random interactions.

3.1 Inclusive Fitness in the Snowdrift and Stag Hunt Games

To rescue the notion of inclusive fitness it is not only necessary to show that it remains technically valid, but also to demonstrate its utility. To that end we show how the notion of inclusive fitness can be used to analyse two simple classes of non-additive games: the snowdrift and stag hunt games. Both games are cooperative dilemmas in that the population would be best served by everyone cooperating, but selection does not always reach this state [9]. In the snowdrift game there is a single stable fixed point with an intermediate level of cooperation. In the stag hunt game, whilst all cooperate is indeed stable, so is all defect, which particular state is reached depends on the initial conditions. The central claim of inclusive fitness is that relatedness facilitates the evolution of cooperation; here we show that this remains true even in these two non-additive games. In the case of the snowdrift game relatedness increases the level of cooperation at the mixed equilibrium. Furthermore, there exists a level of relatedness which is sufficient for the fixation of cooperation. In the case of the stag hunt game relatedness increases the size of the basin of attraction for the cooperative state, and likewise, there is a level of relatedness which is sufficient to increase the basin of attraction to all initial conditions (provided of course that there is a non-zero level of cooperation to begin with).

The snowdrift game is a game in which $S > P$ and $T > R$ (see fig. 1). Even in the absence of relatedness an intermediate level of cooperation is stable. The stable polymorphic equilibrium is given by [14]:

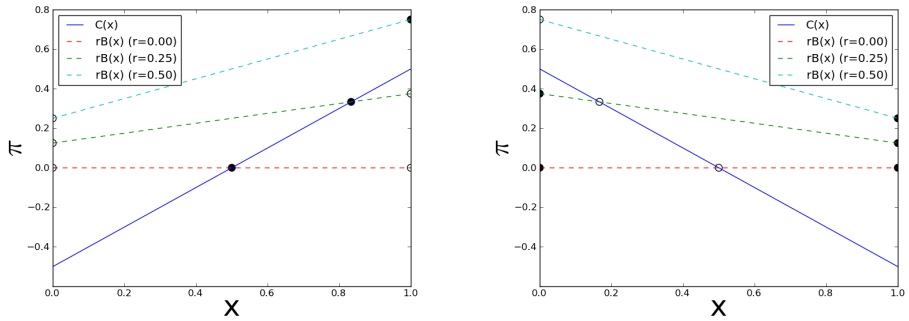
$$x^* = \frac{S-P}{S+T-P-R} \tag{15}$$

This result can be arrived at from our inclusive fitness formalism. A level of cooperation, x, will be stable if $rB(x) = C(x)$. In the well-mixed case $r = 0$ and

thus the fixed point can be found by setting $C(x) = 0$. Recall that $C(x) = x(T-R) + (1-x)(P-S)$. Setting equal to zero and solving for x:

$$\begin{aligned} x(T-R) + (1-x)(P-S) &= 0 && (16) \\ &= x(T-R+S-P) + P - S && (17) \\ x &= \frac{S-P}{S+T-P-R} && (18) \end{aligned}$$

which is equation (15).



(a) Snowdrift Game (b) Stag Hunt Game

Figure 2: Determination of fixed points with frequency dependent costs and benefits. Left: snowdrift game with $(P, R, S, T) = (0, 1, 0.5, 1.5)$ and right: stag hunt game with $(P, R, S, T) = (0, 1, -0.5, 0.5)$. Solid line is $C(x)$ and dashed lines are $rB(x)$ for $r = 0, 0.25$ and 0.5 respectively. A fixed point occurs where the dashed line intercepts the solid line, and furthermore the fixed point is stable if the dashed line crosses from above to below the solid line (in the direction of increasing x). The $r = 0.5$ line does not cross the solid line within the interval which indicates that this level of relatedness is sufficient for cooperation to fixate (in both cases).

One can investigate the effects of relatedness on the snowdrift game by analysing the more general equation: $rB(x) = C(x)$. Substituting our formulae for C and B and rearranging leads to:

$$x^* = \frac{P - rR + rS - S}{(1-r)(P+R-S-T)} \tag{19}$$

this quantity is greater than 1 if $r > {}^{(T-R)}/_{(T-P)}$, which gives us the level of relatedness required for the fixation of cooperation in the snowdrift game.

The stag hunt game occurs when $T < R$ and $S < P$ (see fig. 1). This game also has a fixed point given by equation (15); however, in this case the fixed point is unstable. It therefore defines the basin of attraction for the pure cooperative state.

The effects of increasing relatedness can be studied by looking at the sign of

dx^*/dr.

$$\frac{dx^*}{dr} = \frac{1}{(1-r)^2}\frac{P-R}{P+R-S-T} \tag{20}$$

This is an increasing function of r if: $S+T > P+R$. This is exactly the region of game space above the principle diagonal of figure 1 which includes the snowdrift game.

This situation is represented graphically in figure 2, which shows a gradual increase of cooperation at the stable polymorphic equilibrium in the snowdrift game which reaches 100% for sufficiently high r. For the stag hunt game the figure illustrates an increase of the size of the basin of attraction for the pure cooperative state until the basin covers the whole region for sufficiently high r.

4 Discussion

The inclusive fitness approach to social evolution depends on three terms. Firstly, the coefficient of relatedness which measures the extent to which interactions are correlated between like social strategies. Secondly, cost, which characterises the expected decrease in the number of offspring a certain strategy is likely to incur, and finally benefit, which measures the additional expected number of offspring that a partner of the focal individual can expect on account of that interaction. The first of these three terms has been the one which has attracted the most attention and discussion since the original formulation of inclusive fitness in the 1960s. However, judging by recent debate, the other two terms have caused just as much, if not more, confusion and controversy in the literature. One point, however, which all parties seem to agree with is that for additive games the decomposition is always valid. However, additive games form a vanishing small subset of all possible games and therefore the assumption of additivety is unacceptably restrictive. Furthermore, interesting qualitative phenomena, such as stable polymorphic equlibria, or multiple stable states, can only occur if a game is non-additive. On the face of it this is a fatal blow to the formalism of inclusive fitness. However, this paper has shown that for any possible (pair-wise) game, with any number of strategies there exists an additive payoff matrix for which the direction of selection is the same as the original payoff matrix. Furthermore, we show exactly how to calculate such terms using simple matrix operations. This equivalence comes at a cost, however, which negates some of the simple intuition which inclusive fitness offers. That is that in such a non-additive game the costs and benefits must depend upon the state of the population, and can potentially change from one generation to the next. This may not necessarily be a weakness of any particular formalism itself, but merely a reflection of the fact that non-additive games are inherently more complex systems than additive ones and that any formalism which tackles them must be somewhat less simple and elegant than the simplified version of Hamilton's rule. Nonetheless, we have shown that the central claim of inclusive fitness, namely that relatedness facilitates the evolution of cooperation, remains true. This happens either by increasing the level of cooperation at the stable polymorphic equilibrium (until fixation, for sufficiently

large r) or by increasing the size of the basin of attraction of the cooperative state (likewise to cover the whole interval for sufficiently large r).

Acknowledgements

This work was supported by an EPSRC Doctoral Training Centre grant (EP/G03690X/1). Thanks to Paul Ryan for his useful comments and to Adam Jackson for his useful discussion at the beginning of this work.

References

[1] Ilan Eshel and L. L. Cavalli-Sforza. Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335, 1983.

[2] Jeffrey A Fletcher and Martin Zwick. Unifying the Theories of Inclusive Fitness and Reciprocal Alturism. *The American naturalist*, 168(2):252–262, 2006.

[3] Andy Gardner, Stuart A West, and Geoff Wild. The genetical theory of kin selection. *Journal of evolutionary biology*, 24(5):1020–1043, 2011.

[4] Peter Godfrey-Smith. Varieties of Population Structure and the Levels of Selection. *The British Journal for the Philosophy of Science*, 59(1):25–50, March 2008.

[5] A Grafen. How not to measure inclusive fitness. *Nature*, 1982.

[6] W D Hamilton. The genetical evolution of social behaviour. II. *Journal of theoretical biology*, 7(1):17–52, July 1964.

[7] WD Hamilton. The genetical evolution of social behaviour. I. *Journal of theoretical biology*, 1964.

[8] B. Hölldobler and E. O. Wilson. *The ants*. Harvard University Press, 1990.

[9] Michael W Macy and Andreas Flache. Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences of the United States of America*, 99 Suppl 3:7229–36, May 2002.

[10] John Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, UK, 1982.

[11] R E Michod and D Roze. Cooperation and conflict in the evolution of multicellularity. *Heredity*, 86(Pt 1):1–7, January 2001.

[12] Richard E. Michod and M. J. Sanderson. Behavioral structure and the evolution of cooperation. *Evolution-Essays in honor of John Maynard Smith*, pages 95–104, 1985.

[13] Martin Nowak and Karl Sigmund. The evolution of stochastic strategies in the prisoner's dilemma. *Acta Applicandae Mathematicae*, 20(3):247–265, 1990.

[14] Martin A. Nowak. *Evolutionary Dynamics: Exploring the Equations of Life*. Belknap Press, 2006.

[15] Martin a Nowak, Corina E Tarnita, and Edward O Wilson. The evolution of eusociality. *Nature*, 466(7310):1057–62, August 2010.

[16] Hisashi Ohtsuki. Does synergy rescue the evolution of cooperation? an analysis for homogeneous populations with non-overlapping generations. *Journal of Theoretical Biology*, 307:20–28, 2012.

[17] F. C. Santos, J. M. Pacheco, and T. Lenaerts. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences of the United States of America*, 103(9):3490–4, February 2006.

[18] Elliott Sober. The Evolution of Altruism: Correlation, Cost, and Benefit. *Biology and Philosophy*, 7(2):177–187, 1992.

[19] Joan E Strassmann, Yong Zhu, and David C Queller. Altruism and social cheating in the social amoeba dictyostelium discoideum. *Nature*, 408(6815):965–967, 2000.

[20] Peter D Taylor and Leo B Jonker. Evolutionary stable strategies and game dynamics. *Mathematical biosciences*, 40(1):145–156, 1978.

[21] Matthijs van Veelen. The replicator dynamics with n players and population structure. *Journal of theoretical biology*, 276(1):78–85, May 2011.

[22] E.O. Wilson. *Sociobiology: The New Synthesis*. Belknap Press of Harvard University Press, 2000.