

TI 2014-013/I
Tinbergen Institute Discussion Paper



A Simple Model of Group Selection that cannot be analyzed with Inclusive Fitness

*Matthijs van Veelen*¹

*Shishi Luo*²

*Burton Simon*³

¹ Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute;

² Los Alamos National Laboratory;

³ University of Colorado Denver, United States of America.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: <http://www.dsf.nl/>

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

A simple model of group selection that cannot be analyzed with inclusive fitness

Matthijs van Veelen¹, Shishi Luo² and Burton Simon³

Abstract

A widespread claim in evolutionary theory is that every group selection model can be recast in terms of inclusive fitness. Although there are interesting classes of group selection models for which this is possible, we show that it is not true in general. With a simple set of group selection models, we show two distinct limitations that prevent recasting in terms of inclusive fitness. The first is a limitation across models. We show that if inclusive fitness is to always give the correct prediction, the definition of relatedness needs to change, continuously, along with changes in the parameters of the model. This results in infinitely many different definitions of relatedness - one for every parameter value - which strips relatedness of its meaning. The second limitation is across time. We show that one can find the trajectory for the group selection model by solving a partial differential equation, and that it is mathematically impossible to do this using inclusive fitness.

Keywords

Group selection, inclusive fitness, kin selection, equivalence, social evolution.

¹Department of Economics and Business, University of Amsterdam, and Tinbergen Institute. ²Theoretical Biology and Biophysics (T-6) and Center for Nonlinear Studies, Los Alamos National Laboratory. ³Department of Mathematical and Statistical Sciences, University of Colorado Denver.

Group selection has always been a controversial issue. It is both advocated as an essential ingredient of human evolution (1-2), and described as a superfluous concept, that does not explain any phenomenon we do not already understand with other models (3-5). While there is disagreement concerning the relevance, both sides of the debate have come to agree that group selection models can always be reformulated in terms of inclusive fitness (6-13). The agreement on the equivalence still leaves room to disagree which one of the two is more valuable for understanding how group selection works. One position is that, since both are equivalent, there is no reason to look at group selection models other than through the lens of inclusive fitness (4, 7, 9, 11-13). Another position is that, even though the two are equivalent, there is value in the alternative way of looking at models, that stresses a balance of within-group selection and between-group selection (1, 2, 8, 10).

Because the “mathematical equivalence” is such a central point of consensus, one would expect it to be a well-defined mathematical statement with a mathematical proof. There is, however, no such theorem in the literature (see also 6, 14). As a consequence it remains unclear what “mathematical equivalence” means here, and if the claim is correct. In this paper, we will present a class of group selection models that allows us to explore what it could mean for group selection models and inclusive fitness to be equivalent. It also uncovers two different kinds of limitations. The limitations imply that group selection and inclusive fitness are not equivalent in general, even though they can be under certain assumptions.

The model is a generalization of (15). Individuals find themselves in groups of equal size, and can be one of two types; defectors and cooperators. Individual reproduction happens at different rates; cooperators reproduce at rate 1, defectors reproduce at rate $1 + s$. Every individual reproduction event induces a death event; if one individual reproduces, one random individual from the same group is chosen for elimination, thereby keeping group sizes

constant. Entire groups also reproduce, and when they do, they produce a daughter group with the exact same proportion of cooperators and defectors. The rate at which this occurs depends on the fraction of cooperators in the group; a group of size n with i cooperators in it reproduces at rate $1 + u \left(\frac{i}{n}\right)^\alpha$. That implies that all-defector groups reproduce at rate 1, and all-cooperator groups reproduce at rate $1 + u$. Every group reproduction event induces a group death event; if a group reproduces, a random group is chosen for elimination (see Figure 1).

Being a cooperator therefore comes at a cost - it reduces the reproduction rate of the individual by s - but it has a benefit for all group members, including itself, through an increase in the rate at which the group as a whole reproduces. The baseline reproductive rates for individuals and groups are set to unity for simplicity. The Appendix also covers the more general case where the baseline reproduction rates of individuals and groups are allowed to differ. A setup that allows for groups of different sizes and a variety of group level events, is also possible (16-18).

If we take a limit, where group size and number of groups go to infinity, then the dynamics simplify to a partial differential equation (PDE). In the limit, the fraction of cooperators in the group becomes a continuous variable, x , which ranges from 0 to 1. A population at time t is a density $\mu_t(x)$ that reflects the relative frequencies of groups with different shares of cooperators. The PDE describes how this density changes over time due to the process of group selection (see Figure 2).

$$\frac{d}{dt}\mu_t(x) = s\frac{d}{dx}[x(1-x)\mu_t(x)] + u\mu_t(x)\left[x^\alpha - \int_0^1 y^\alpha\mu_t(y)dy\right]$$

The first term on the right hand side describes how individual groups see their fraction of cooperators x go down. The second term describes how

high- x -groups increase and low- x -groups decrease in frequency. Between the square brackets, the term x^α reflects the increase in group reproduction rate of an x -group that is caused by the cooperators in that group. The integral $\int_0^1 y^\alpha \mu_t(y) dy$ does not depend on x ; it reflects a uniform increase in group death rate that is caused by the fact that cooperators in the entire population increase the growth rates of their groups. Group reproduction events are balanced by group deaths, so the increase in group death intensity must match the average increase in group reproduction - which is $\int_0^1 y^\alpha \mu_t(y) dy$. The PDE is derived in the Appendix.

In order to illustrate how inclusive fitness *can* match the prediction of a group selection model, we first consider the case where $\alpha = 1$. Let $M_k(t) = \int_0^1 x^k \mu_t(x) dx$ be the k th moment of the density μ_t . We are interested in the rate of change of the first moment, which is the fraction of cooperators in the overall population. The formula is obtained by integrating the right hand side of the PDE (see the Appendix). For conciseness we will suppress the dependence on t in the notation, assuming that it is clear that M_1 and M_2 change with t .

$$\frac{d}{dt}M_1 = s(M_2 - M_1) + u(M_2 - (M_1)^2) \quad (1)$$

If the time derivative $\frac{d}{dt}M_1$ is larger than 0, then the frequency of cooperators is increasing; if it is smaller than 0, the frequency is decreasing. The right hand side of the equation is composed of two terms. The first term is negative, or 0, and measures the effect of within group selection, which works against cooperators. If all groups are at $x = 0.5$, this term is at its minimum - the frequency of cooperators goes down the fastest - and if all groups are either close to $x = 0$ or close to $x = 1$, then this term is almost 0, and the frequency of cooperators hardly changes. The second term is u times the variance of the density μ_t . This term is positive, or 0, and measures the effect of between

group selection, which works against defectors. This formula thereby reflects an intuition that many have for group selection models; whether or not the share of cooperators increases depends on which of the two is larger.

If we rewrite this formula, we arrive at an equally insightful and intuitive expression, this one in terms of inclusive fitness.

$$\begin{aligned} \frac{d}{dt}M_1 &= M_1(1 - M_1) \left[\left(\frac{M_2 - (M_1)^2}{M_1 - (M_1)^2} \right) (s + u) - s \right] \\ &= M_1(1 - M_1)(rb - c) \end{aligned}$$

The $M_1(1 - M_1)$ term scales the speed of selection, and $rb - c$ is the most natural definition of inclusive fitness in this setting. Whether or not the frequency of cooperators goes up is determined by whether or not inclusive fitness is larger than 0. Defining relatedness as $r = \frac{M_2 - (M_1)^2}{M_1 - (M_1)^2}$ is in line with standard definitions in the literature (12, 13, 19-21; see also the Appendix). Moreover, s is indeed the cost of cooperating to the individual; it is the reduction in individual reproduction rate that results from being a cooperator instead of a defector. The aggregate benefits to all other group members together are $b = s + u$. Reducing individual reproduction by s implies reducing the aggregate death rate of the other group members by the same amount, because every individual reproduction is balanced by another individual being eliminated. Moreover, in a group of size n , the contribution of one individual cooperator to the group reproduction rate is u times $1/n$. Since this increase is enjoyed by all $n - 1$ other group members, the aggregate benefits to all other group members is $u \cdot \frac{n-1}{n}$, which, in the limit of infinitely large groups, is u . Hence total aggregate benefits are $s + u$. The benefit to the individual itself through increased group reproduction is 0 in infinitely large groups.

How does this change if we allow for general $\alpha > 0$? First we compute the overall change in the frequency of cooperators (see the Appendix).

$$\frac{d}{dt}M_1 = s(M_2 - M_1) + u(M_{\alpha+1} - M_1M_\alpha)$$

If we take $\alpha = 1$, we recover equation (1).

Trying to recast this in terms of inclusive fitness, we first observe that with $\alpha \neq 1$ it is possible that two different population states have the same relatedness, and yet in one of them the frequency of cooperators is going up, and in the other the frequency of cooperators is going down (see Figure 3). That implies that there are no fixed parameters b and c such that the frequency of cooperators goes up if, and only if, $rb > c$. But also if we allow benefits and costs to depend on the current population state, it is still not the case that inclusive fitness accurately matches the direction of selection. In this model, the costs of cooperation are independent of the population state, and should therefore be constant; changing from a defector to a cooperator always means reducing individual reproduction rate by s . The benefits, however, depend on the composition of the group an individual is in. Therefore, also the average benefits dispensed by cooperators will no longer be constant, as it will depend on the current distribution. The Appendix contains a derivation, from which it follows that average benefits are $s + u\alpha\frac{M_\alpha}{M_1}$, which depends on the current population density through M_α and M_1 . However, even if we use the average costs and benefits, it remains the case that the signs of $rb - c$ and the change in cooperator frequency are sometimes opposing (see the Appendix).

It has been argued that the appropriate approach is then to adjust the definition of relatedness, such that inclusive fitness does give the right prediction (22-25). If we follow this approach, as we do in the Appendix, then we find that the relatedness that would be needed is:

$$r = \frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \frac{s + u}{s + u \alpha \frac{M_\alpha}{M_1}}$$

With this approach, the formula for relatedness comes to depend, not only on the current population structure, but also on all parameters that describe the fitness effects. This implies that every other choice for s , u and α would require a different definition of relatedness. Relatedness then ceases to be a property of the population structure that has an interpretation of its own, and that can be measured independent of the fitness effects in the model. With the definition of relatedness continuously changing, every case gets its own “rule”, which disqualifies it from being one (26).

For the second limitation, we return to the case where $\alpha = 1$. For that case, we have already shown that at any point in time, the direction of selection matches the sign of inclusive fitness. However, if we focus on the trajectory as a whole, we see that relatedness is not constant. As the trajectory unfolds, the density changes, and, since both the frequency of cooperators M_1 and relatedness r are characteristics of that density, they change too (see Figure 4). If we have the solution to the PDE, then we also know how r , and M_1 , change over time. Without already having the solution, however, this is not possible. Relatedness r is a function of M_1 and M_2 - the first and second moment of the density function - but in order to determine how r changes over the trajectory, we need all moments of the distribution; relatedness alone does not contain enough information. Only once we already have the entire solution, can we compute r and M_1 along the trajectory, and conclude that $\frac{d}{dt} M_1$ is positive if and only if $rb - c > 0$. Inclusive fitness therefore does not offer an alternative way to find the solution; we need to already have the solution in order to determine what inclusive fitness is along the trajectory.

Both these limitations imply that it is not the case that everything one can do with a group selection model can also be done with inclusive fitness. That,

however, does not mean that there are no assumptions under which there is in fact equivalence. A combination of conditions under which group selection and inclusive fitness are equivalent is that 1) relatedness is exogenously given, and constant, and 2) groups play a public goods game with additive fitness, or “generalized equal gains from switching” (27). Assuming a process where mutations cause small changes in phenotype, in combination with a fitness function that is differentiable is one way to get both a fixed relatedness and equal gains from switching (9). Whether or not such assumptions are realistic is an empirical question, which our theoretical findings cannot answer. They do however show that group selection and inclusive fitness are not mathematically equivalent. This implies that there can be value in exploring group selection models beyond what one can do with inclusive fitness. Richer models of group selection (15-18) might for instance describe under what circumstances we should expect selection to move up one level entirely and when it remains a balance between two levels. They might also identify conditions under which a diversity of social attitudes is stable, as it is in this model, where the stable steady state is a distribution in which both types are present (see the Appendix). In many applications the goal is also not just to find out whether the fraction of cooperators will increase at a single point in time, but to know what the ultimate fraction cooperators will be, or the change of the fraction of cooperators over a period of time. Models that feature both group and individual level events explicitly, like the ones here and in (ref 14-18) can tell us what we need to know about the evolutionary dynamics. The fact that not every group selection model can be analyzed with inclusive fitness therefore opens up a range of possibilities that need to be explored, and allows us to pose and answer questions that do not fit the format of inclusive fitness.

Acknowledgements SL was supported by the Duke Stern Fellowship and the US Department of Energy through the LANL/LDRD Program.

References

- [1] Sober E, Wilson DS (1998) *Unto Others; the evolution and psychology of unselfish behavior* (Harvard University Press, Cambridge, MA).
- [2] Wilson DS., Wilson EO (2007) Rethinking the theoretical foundations of socio-biology. *Q Rev Biol.* 82:327-348.
- [3] Williams GC (1966) *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought* (Princeton Univ. Press, Princeton).
- [4] Dawkins R (1989) *The Extended Phenotype* (Oxford Univ. Press, Oxford).
- [5] Pinker S (2012) The false allure of group selection, *Edge*, June 21, 2012, <http://edge.org/conversation/the-false-allure-of-group-selection>.
- [6] Queller DC (1992) Quantitative genetics, inclusive fitness, and group selection. *Am Nat* 139:540-558.
- [7] Lehmann L, Keller L, West SA, Roze D (2007) Group selection and kin selection: two concepts but one process. *Proc Natl Acad Sci USA* 104:6736-6739.
- [8] Hölldobler B, Wilson EO (2009) *The Superorganism: The Beauty, Elegance, and Strangeness of Insect Societies* (W. W. Norton, New York).
- [9] Wild G, Gardner A, West SA (2009) Adaptation and the evolution of parasite virulence in a connected world. *Nature* 459:983–986.
- [10] Wade MJ et al. (2010). Multilevel and kin selection in a connected world. *Nature* 463:E8-E9.

- [11] Lion S, Jansen VAA, Day T (2011). Evolution in structured populations: beyond the kin versus group debate. *Trends Ecol Evol* 26:193-201.
- [12] Marshall JAR (2011) Group selection and kin selection: formally equivalent approaches. *Trends Ecol Evol* 26:325-332.
- [13] Gardner A, West SA, Wild G (2011) The genetical theory of kin selection. *J Evol Biol* 24:1020-1043.
- [14] van Veelen M, García J, Sabelis MW, Egas M (2012) Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics *J Theor Biol* 299:64–80.
- [15] Luo SA (2014) A unifying framework reveals key properties of multilevel selection. *J Theor Biol* 341:41–52
- [16] Simon B (2010) A dynamical model of two-level selection. *Evol Ecol Res* 12:555–588.
- [17] Simon B, Fletcher JA, Doebeli M (2012) Hamilton’s rule in multi-level selection models, *J Theor Biol* 299:55–63.
- [18] Simon B, Fletcher JA, Doebeli M (2013) Towards a general theory of group selection, *Evolution* 67:1561–1572.
- [19] Orlove MJ, Wood CL (1978) Coefficients of relationship and coefficients of relatedness in kin selection: a covariance form for the RHO formula. *J Theor Biol* 73:679–686.
- [20] Eshel I, Cavalli-Sforza LL (1982) Assortment of encounters and evolution of cooperativeness. *Proc Natl Acad Sci USA* 79:1331–1335.
- [21] Queller DC (1992) A general model for kin selection. *Evolution* 46:376–380.

- [22] Grafen A (2006) Optimization of inclusive fitness. *J Theor Biol* 238:541–563.
- [23] Grafen A (2007) Detecting kin selection at work using inclusive fitness. *Proc R Soc B* 274:713-719.
- [24] Grafen A (2009) Formalizing Darwinism and inclusive fitness theory *Phil. Trans. R. Soc. B* 364:3135–3141.
- [25] Marshall JAR (2011) Queller’s rule ok: Comment on van Veelen ‘when inclusive fitness is right and when it can be wrong’. *J Theor Biol* 270:185-188.
- [26] Van Veelen M (2011) A rule is not a rule if it changes from case to case (a reply to Marshall’s comment). *J Theor Biol* 270:189-195.
- [27] Van Veelen M (2011) The replicator dynamics with n player games and population structure. *J. Theor. Biol.* 276:78-85.

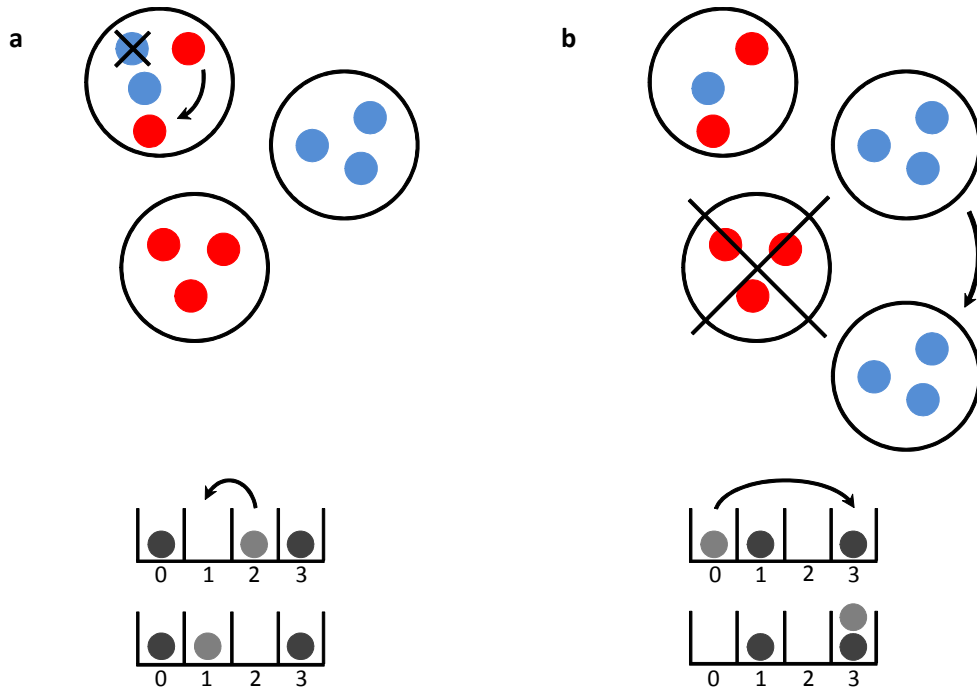


Fig. 1. Panel (a) depicts an individual level event, panel (b) a group level event. Cooperators (blue) reproduce individually at intensity 1, defectors (red) at intensity $1 + s$. When an individual reproduces, a random individual is chosen to die. Groups reproduce at intensity $1 + u \left(\frac{i}{n}\right)^\alpha$, where i is the number of cooperators in a group, and n the group size. When a group reproduces, a random group is chosen to die. The effect of the events on the composition of the population is visualized below each panel; they change how many groups there are with 0, 1, 2 and 3 cooperators, respectively. In the limit of the group size going to infinity, groups are characterized by the fraction x of cooperators, and as the number of groups goes to infinity, the population state becomes a density over $(0, 1)$, that describes how many groups there are for different cooperator frequencies x .

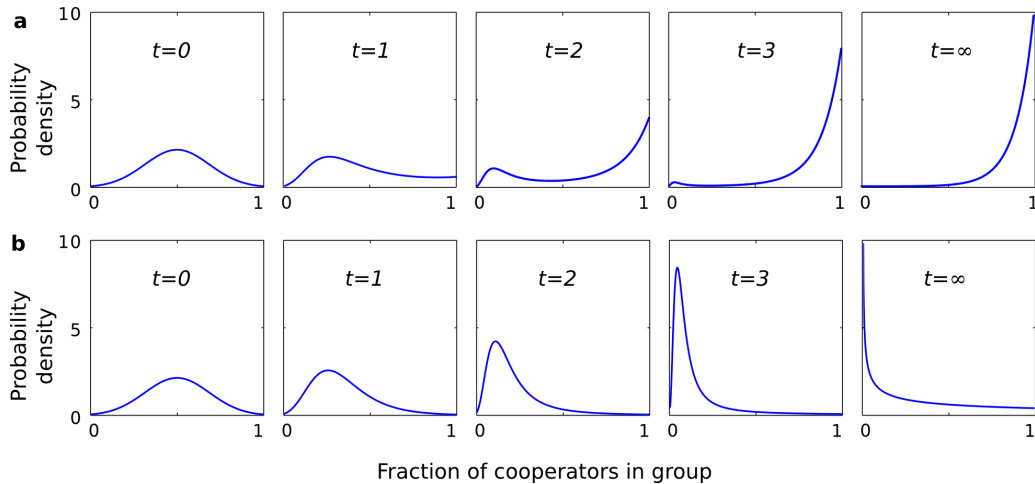


Fig. 2. A density describes the relative frequencies of groups with different fractions x of cooperators, where x varies between 0 and 1. The PDE describes how these densities change over time. The changes are caused by two effects. The first is that individual reproduction makes all groups become less cooperative - which makes the "humps" in the distributions move to the left in both examples. The second is that cooperative groups as a whole reproduce faster - which results in the right end of the distributions being lifted up. As time elapses, the densities change, until the two effects balance at a stable steady state, described by the equilibrium distribution. The upper series (a) has $\alpha = 3$, $s = 1$ and $u = 4$, the lower series (b) has $\alpha = 1$, $s = 1$ and $u = 1.5$. For both the initial distribution is proportional to a truncated normal distribution with mean 0.5 and variance 0.035. In both cases the densities converge to a steady state density in which all group types are represented, but in the first case there are more cooperators than defectors, overall, and in the second there are more defectors than cooperators.

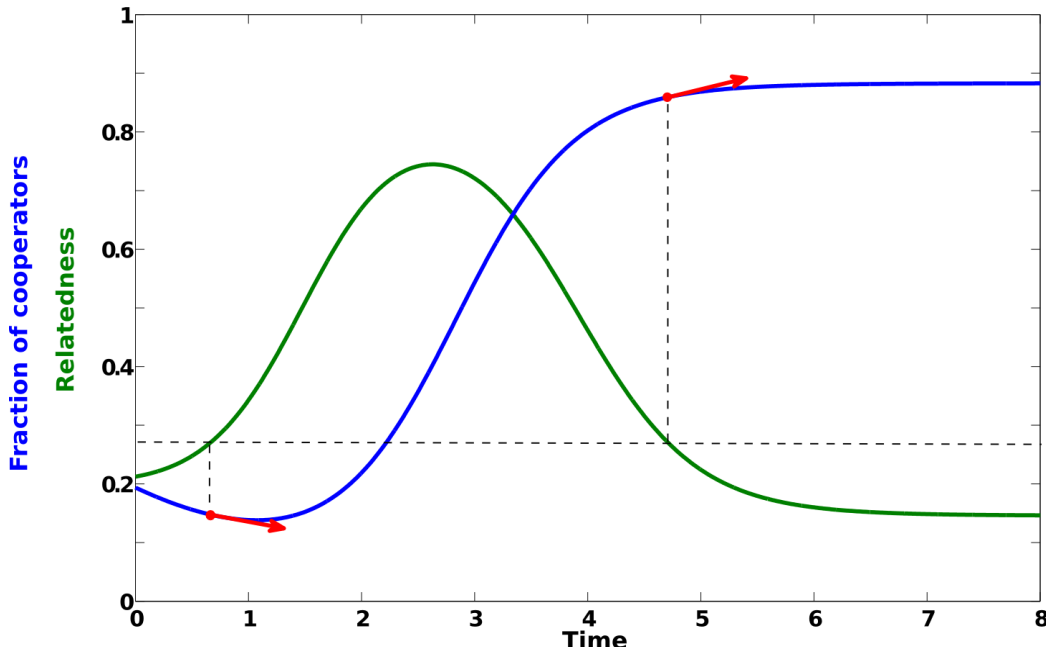


Fig. 3. The blue line tracks the fraction of cooperators in the population, and the green line tracks relatedness, when $\alpha = 4$, $s = 1$, $u = 3$ and initial distribution $\mu_0(x) = 5e^{-5x}/(1 - e^{-5})$. Relatedness is the same at the two points in time indicated by the arrows. Yet at the first one the fraction of cooperators decreases, and at the other it increases. This discrepancy remains if we correct for average costs and benefits not being the same at different moments in time. If we allow costs and benefits to depend on the current density, and use average costs for c and average benefits for b , then the direction of selection is still not described by the sign of $rb - c$ (see the Appendix). Also, the steady state that the population converges to does not satisfy $rb - c = 0$, with average benefits b and average costs c .

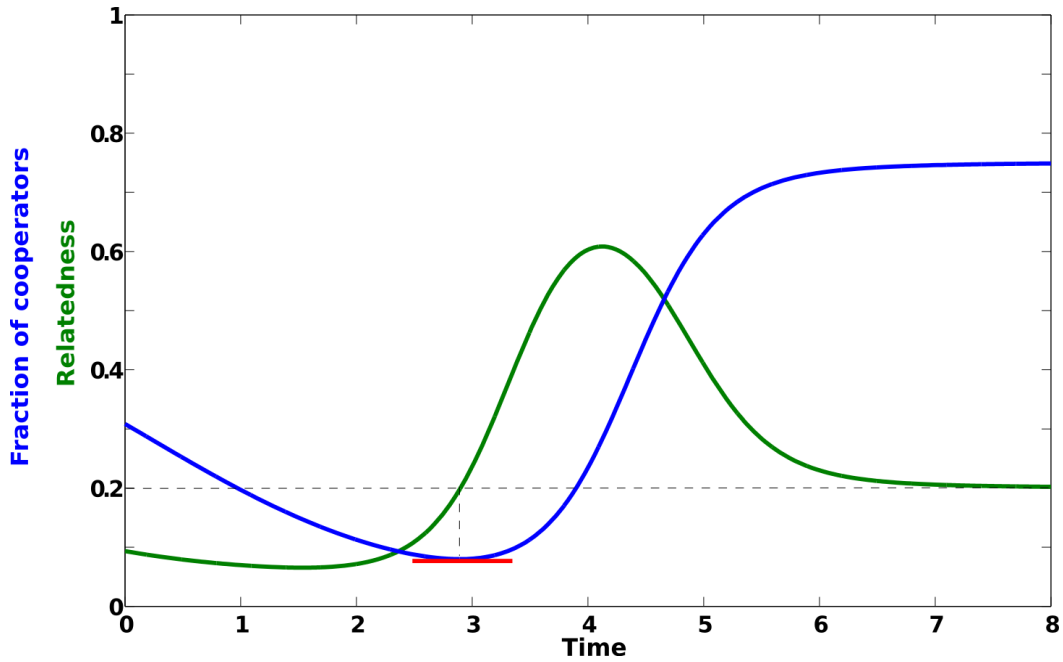


Fig. 4. The blue line tracks the fraction of cooperators in the population, and the green line tracks relatedness, when $\alpha = 1$, $s = 1$, $u = 4$ and the initial distribution $\mu_0(x)$ is proportional to a truncated normal distribution with mean 0.3 and variance 0.0225. With $\alpha = 1$, the direction of selection is accurately described by the sign of $rb - c$. The benefits-to-costs ratio is constant at $b/c = s/(s + u) = 0.2$. The fraction of cooperators increases whenever $r > 0.2$, and decreases when $r < 0.2$. However, knowing the relatedness at time $t = 0$ is not enough to determine what relatedness will be at $t > 0$. For that we need to solve the partial differential equation first. Since the solution itself already describes how the frequency of cooperators changes over time, inclusive fitness does not offer a route to finding the answer; determining what the inclusive fitness is requires that we already have the answer.

Appendix to:
A simple model of group selection that cannot be
analyzed with inclusive fitness

Matthijs van Veelen^{1,2}, Shishi Luo³, Burton Simon⁴

¹Department of Economics and Business, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, ²Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts, 02138, USA. ³Theoretical Biology and Biophysics (T-6) and Center for Nonlinear Studies, MS B258, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ⁴Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, Colorado 80202, USA.

1 Introduction

The appendix begins with the technical results. In Section 2 the partial differential equation is derived. This is the most technical section. Later sections use the PDE, and can also be understood without understanding the details of its derivation. In Section 3 we show that inclusive fitness matches the direction of selection for $\alpha = 1$. In Section 4 we derive that, if inclusive fitness is to match the direction of selection for $\alpha > 0$, relatedness will have to depend on all parameters that describe the fitness effects of cooperation. In Section 5 we show why relatedness only is not enough information about the current population state for finding a trajectory. In Section 6 we allow for individual and group baseline reproduction rates to differ and we find stable steady states. After the technical results, in Section 7, we discuss the results in some more detail than in the main text, which will be useful because of the polarized nature of the debate and the sensitivity of both group selection and inclusive fitness.

2 Derivation of the equation for the dynamics

The stochastic model described in the main text is similar to the ball-and-urn process described in Luo (2013). Let n be the number of individuals in each group and m the number of groups in the population. The main differences between Luo (2013) and the derivation here are that in the model of the main text, a group's reproductive rate is of a more general form, $1 + u \cdot (\frac{i}{n})^\alpha$, and w is fixed at $w = 1$. In fact, here in the SI we will assume an even more general form, $1 + f(\frac{i}{n})$ for some function f , to demonstrate the full generality of this model. Letting $\mu_t^{m,n}$ denote the stochastic model described in the main text, we can take the limit as $m, n \rightarrow \infty$ in a similar manner to the limit in Luo (2013). This gives a law of large numbers, or a fluid limit, for the dynamics. The calculations also bear close resemblance to standard derivations of diffusion limits for standard Moran processes (Karlin and Taylor, 1981, Durrett, 2008).

To calculate the infinitesimal mean, or drift, of $\mu_t^{m,n}$, note that:

$$\begin{aligned}
 & \mathbb{E}[\mu_{t+\Delta t}^{m,n}(\frac{i}{n}) - \mu_t^{m,n}(\frac{i}{n})] \\
 = & -\frac{1}{m} \left\{ m\mu_t^{m,n}(\frac{i}{n})i \left(1 - \frac{i}{n}\right) (2 + s) + m\mu_t^{m,n}(\frac{i}{n}) \sum_{j \neq i} \mu_t^{m,n}(\frac{j}{n})(1 + f(\frac{j}{n})) \right\} \Delta t \\
 & + \frac{1}{m} \left\{ m\mu_t^{m,n}(\frac{i+1}{n})(n - (i + 1)) (1 + s) \left(\frac{i+1}{n}\right) + m\mu_t^{m,n}(\frac{i-1}{n})(i - 1) \left(1 - \frac{i-1}{n}\right) \right. \\
 & \left. + m\mu_t^{m,n}(\frac{i}{n})(1 + f(\frac{i}{n}))(1 - \mu_t^{m,n}(\frac{i}{n})) \right\} \Delta t + o(\Delta t)
 \end{aligned} \tag{1}$$

The first bracketed term represents the events that decrease the fraction of groups with i cooperators by $1/m$. This can happen if, in a group with i cooperators, a cooperator replicates and a defector dies or a defector dies and a cooperator replicates (first term, inside the brackets). Or, if a group with i cooperators is selected to die when another group, with a different number

of cooperators, replicates (second term, inside the brackets). The second bracketed term analogously represents the events that increase the fraction of groups with i cooperators by $1/m$. A detailed explanation of these terms can be found in Luo (2013). This standard approach for deriving a diffusion limit uses the fact that the probability of an event occurring in a small time interval, Δt , is equal to the rate that the event occurs multiplied by Δt . The $o(\Delta t)$ term represents higher order terms corresponding to compound events (more than one event occurring in the interval Δt), which can be ignored in the limit as $\Delta t \rightarrow 0$. Thus, after some rearranging of (1), we find

$$\begin{aligned} & \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{E} [\mu_{t+\Delta t}^{m,n} \left(\frac{i}{n} \right) - \mu_t^{m,n} \left(\frac{i}{n} \right)] = \\ & = \frac{1}{n} D_2 \left(\mu_t^{m,n} \left(\frac{i}{n} \right) \frac{i}{n} \left(1 - \frac{i}{n} \right) \right) + s D_1^+ \left(\mu_t^{m,n} \left(\frac{i}{n} \right) \frac{i}{n} \left(1 - \frac{i}{n} \right) \right) \\ & \quad + \mu_t^{m,n} \left(\frac{i}{n} \right) \left[f \left(\frac{i}{n} \right) - \sum_{j=0}^n \mu_t^{m,n} \left(\frac{j}{n} \right) f \left(\frac{j}{n} \right) \right] \end{aligned}$$

where

$$D_1^+ \left(g \left(\frac{j}{n} \right) \right) = \frac{g \left(\frac{j+1}{n} \right) - g \left(\frac{j}{n} \right)}{1/n} \quad j < n$$

is a first-order difference quotient and

$$D_2 \left(g \left(\frac{j}{n} \right) \right) = \frac{g \left(\frac{j+1}{n} \right) - 2g \left(\frac{j}{n} \right) + g \left(\frac{j-1}{n} \right)}{1/n^2} \quad 0 < j < n$$

is the second-order difference quotient. On calculation of the infinitesimal variance, we find that:

$$\begin{aligned}
& \mathbb{E} \left[\mu_{t+\Delta t}^{m,n} \left(\frac{i}{n} \right) - \mu_t^{m,n} \left(\frac{i}{n} \right) \right]^2 \\
&= \frac{1}{m^2} \left\{ m \mu_t^{m,n} \left(\frac{i}{n} \right) \frac{i}{n} \left(1 - \frac{i}{n} \right) (2 + s) + m \mu_t^{m,n} \left(\frac{i}{n} \right) \sum_{j \neq i} \mu_t^{m,n} \left(\frac{j}{n} \right) \frac{j}{n} (1 + f \left(\frac{j}{n} \right)) \right. \\
&\quad + m \mu_t^{m,n} \left(\frac{i+1}{n} \right) \left(\frac{i+1}{n} \right) \left(1 - \frac{i+1}{n} \right) (1 + s) + m \mu_t^{m,n} \left(\frac{i-1}{n} \right) \left(\frac{i-1}{n} \right) \left(1 - \frac{i-1}{n} \right) \\
&\quad \left. + m \mu_t^{m,n} \left(\frac{i}{n} \right) (1 + f \left(\frac{i}{n} \right)) \left(1 - \mu_t^{m,n} \left(\frac{i}{n} \right) \right) \right\} \Delta t + o(\Delta t)
\end{aligned}$$

Thus:

$$\lim_{m \rightarrow \infty, \Delta t \rightarrow 0} \frac{1}{\Delta t} \mathbb{E} \left[\mu_{t+\Delta t}^{m,n} \left(\frac{i}{n} \right) - \mu_t^{m,n} \left(\frac{i}{n} \right) \right]^2 = 0 \quad \text{for } i = 0, \dots, n$$

Since the infinitesimal variance is zero, in the infinite population limit, the time evolution of the population is described deterministically by the equation for the infinitesimal mean (1). Taking the limit as $n, m \rightarrow \infty$, we obtain

$$\frac{\partial}{\partial t} \mu_t(x) = s \frac{\partial}{\partial x} [x(1-x)\mu_t] + \mu_t(x) \left[f(x) - \int_0^1 f(y) \mu_t(y) dy \right]$$

Taking $f(x) = u \cdot x^\alpha$, we obtain the PDE for the system described in the main text.

3 Inclusive fitness with $\alpha = 1$

If we take $f(x) = ux$ for the function that describes the increase in group reproduction due to cooperators, then the PDE is:

$$\frac{d}{dt}\mu_t(x) = s\frac{d}{dx}[x(1-x)\mu_t(x)] + u\mu_t(x)\left[x - \int_0^1 y\mu_t(y) dy\right]$$

We now want to describe how the frequency of cooperators in the population as a whole changes. The frequency of cooperators at time t is $\int_0^1 x\mu_t(x) dx$, the first moment of the density μ_t , also denoted as M_1 . We are interested in the time derivative $\frac{d}{dt}\int_0^1 x\mu_t(x) dx = \frac{d}{dt}M_1$ of the first moment. Since the PDE holds for all x , we can multiply left and right by x and integrate from 0 to 1. This will give us an expression for $\frac{d}{dt}M_1$ since $\frac{d}{dt}\int_0^1 x\mu_t(x) dx = \int_0^1 x\frac{d}{dt}\mu_t(x) dx$.

First we consider the first term of the right hand side of the PDE, and use integration by parts.

$$\begin{aligned} s\int_0^1 x\frac{d}{dx}[x(1-x)\mu_t(x)] dx &= s\left(\left[x^2(1-x)\mu_t(x)\right]_0^1 - \int_0^1 [x(1-x)\mu_t(x)] dx\right) = \\ &= -s\int_0^1 [x(1-x)\mu_t(x)] dx = \\ &= s\left(\int_0^1 x^2\mu_t(x) dx - \int_0^1 x\mu_t(x) dx\right) \end{aligned}$$

Then we integrate the second term.

$$u\int_0^1 x\mu_t(x)\left[x - \int_0^1 y\mu_t(y) dy\right] dx = u\left(\int_0^1 x^2\mu_t(x) dx - \left(\int_0^1 x\mu_t(x) dx\right)^2\right)$$

Together that makes:

$$\frac{d}{dt} \int_0^1 x \mu_t(x) dx = s \left(\int_0^1 x^2 \mu_t(x) dx - \int_0^1 x \mu_t(x) dx \right) + u \left(\int_0^1 x^2 \mu_t(x) dx - \left(\int_0^1 x \mu_t(x) dx \right)^2 \right)$$

This can be written in more concise notation as

$$\frac{d}{dt} M_1 = s (M_2 - M_1) + u (M_2 - (M_1)^2),$$

where M_1 and M_2 are short for the first and second moment of $\mu_t(x)$. Note that $M_2 - M_1 \leq 0$ and $M_2 - (M_1)^2 \geq 0$.

If we rearrange the derivative of the first moment, we get the following:

$$\begin{aligned} \frac{d}{dt} M_1 &= s (M_2 - M_1) + u (M_2 - (M_1)^2) \\ &= s M_1 (M_1 - 1) + (s + u) (M_2 - (M_1)^2) \\ &= M_1 (1 - M_1) \left[\left(\frac{M_2 - (M_1)^2}{M_1 - (M_1)^2} \right) (s + u) - s \right] \end{aligned}$$

With $r = \frac{M_2 - (M_1)^2}{M_1 - (M_1)^2}$, $c = s$ and $b = s + u$, this becomes an expression that implies that the direction of selection follows from the sign of $rb - c$, and that the speed of selection is scaled by $M_1 (1 - M_1)$.

$$\frac{d}{dt} M_1 = M_1 (1 - M_1) (rb - c)$$

Below we will verify that r , c and b are indeed the proper relatedness, costs and benefits.

3.1 Relatedness

The first term between square brackets matches the most common definitions of relatedness. One definition, which serves as a measure of assortment in general, is $r = \mathbb{P}(C|C) - \mathbb{P}(C|D)$. In this setting, with infinitely large groups, these probabilities can be described with the following chance experiment. First draw a group, where the probabilities for the type of group that is drawn are described by the density μ_t . Then draw two individuals from that group; both times a cooperator is drawn with probability x . Relatedness is then defined as a difference in conditional probabilities, where $\mathbb{P}(C|C)$ is the probability that the second is a cooperator, given that the first is a cooperator, and $\mathbb{P}(C|D)$ is the probability that the second is a cooperator, given that the first is a defector (see for instance Bergstrom, 2003, Van Veelen 2009, 2011a,b and Bowles & Gintis, 2011). This is also consistent with the definition of assortment in Eshel & Cavalli-Sforza (1982) and of relatedness in Grafen (1985).

With this definition, we find:

$$\begin{aligned} r &= \mathbb{P}(C|C) - \mathbb{P}(C|D) = \frac{\int_0^1 x^2 \mu_t(x) dx}{\int_0^1 x \mu_t(x) dx} - \frac{\int_0^1 (1-x) x \mu_t(x) dx}{\int_0^1 (1-x) \mu_t(x) dx} \\ &= \frac{\int_0^1 x^2 \mu_t(x) dx}{\int_0^1 x \mu_t(x) dx} - \frac{\int_0^1 x \mu_t(x) dx - \int_0^1 x^2 \mu_t(x) dx}{1 - \int_0^1 x \mu_t(x) dx} \\ &= \frac{M_2}{M_1} - \frac{M_1 - M_2}{1 - M_1} = \frac{M_2 - (M_1)^2}{M_1 - (M_1)^2} \end{aligned}$$

An equivalent definition of relatedness is $r = \frac{Cov(X,Y)}{Var(X)}$; see for instance Orlove & Wood (1978), Queller (1992) and Gardner, West & Wild (2011). If we think of the same experiment as described above, we can define random variable X to be 1 if the first individual drawn is a cooperator and 0 if it is a defector, and random variable Y to be 1 if the second individual is a cooperator and 0 if it is a defector. This definition of relatedness then amounts to the same:

$$r = \frac{Cov(X, Y)}{Var(X)} = \frac{\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]}{\mathbb{E}[X^2] - \mathbb{E}^2[X]} = \frac{M_2 - (M_1)^2}{M_1 - (M_1)^2}$$

The reason why $\mathbb{E}[XY] = M_2$ is that X times Y is 1 only if both X and Y are 1, and 0 if not. In an x -group, X is 1 with probability x and Y is too, so $\mathbb{E}[XY] = \int_0^1 (x \cdot x) \mu_t(x) dx = M_2$.

The reason why $\mathbb{E}[X^2] = M_1$ is that X^2 is only 1 if X is 1, and 0 if not. In an x -group, that happens with probability x , so $\mathbb{E}[X^2] = \int_0^1 x \mu_t(x) dx = M_1$.

3.2 Costs and benefits

The difference between the individual reproduction rates of cooperators and defectors is s , which therefore is indeed the cost of cooperating to the individual. For every individual reproduction, another individual in the same group is eliminated, hence reducing individual reproduction by s implies reducing the aggregate death rate of the other group members by the same amount.

In finite groups, the marginal effect on the group reproduction rate is u times one over the group size, but the benefits are shared by everyone in the group, and therefore the aggregate benefits to everyone else in the group is $\frac{n-1}{n} \cdot u$, where n is the group size, and the group benefits to itself are $\frac{1}{n} \cdot u$. With $n \rightarrow \infty$, the benefits to self go to 0, and the aggregate benefits to the other group members go to u . Total aggregate benefits therefore are $s + u$.

4 Inclusive fitness with general $\alpha > 0$

If we take $f(x) = ux^\alpha$ for the function that describes the increase in group reproduction due to cooperators, then the PDE becomes:

$$\frac{d}{dt}\mu_t(x) = s\frac{d}{dx}[x(1-x)\mu_t(x)] + u\mu_t(x)\left[x^\alpha - \int_0^1 y^\alpha\mu_t(y)dy\right] \quad (2)$$

We still are interested in the change in the frequency of cooperators, and therefore we compute $\frac{d}{dt}\int_0^1 x\mu_t(x)dx = \frac{d}{dt}M_1$, the time derivative of the first moment, also for the general case. The first term on the right hand side of the PDE is still the same. The integral for the first term therefore remains

$$s\int_0^1 x\frac{d}{dx}[x(1-x)\mu_t(x)]dx = s\left(\int_0^1 x^2\mu_t(x)dx - \int_0^1 x\mu_t(x)dx\right)$$

The integral over the second term becomes

$$u\int_0^1 x\mu_t(x)\left[x^\alpha - \int_0^1 y^\alpha\mu_t(y)dy\right]dx = u\left(\int_0^1 x^{\alpha+1}\mu_t(x)dx - \left(\int_0^1 x\mu_t(x)dx\right)\left(\int_0^1 x^\alpha\mu_t(x)dx\right)\right)$$

Together that makes

$$\begin{aligned} \frac{d}{dt}\int_0^1 x\mu_t(x)dx &= s\left(\int_0^1 x^2\mu_t(x)dx - \int_0^1 x\mu_t(x)dx\right) + \\ &\quad u\left(\int_0^1 x^{\alpha+1}\mu_t(x)dx - \left(\int_0^1 x\mu_t(x)dx\right)\left(\int_0^1 x^\alpha\mu_t(x)dx\right)\right) \\ \frac{d}{dt}M_1 &= s(M_2 - M_1) + u(M_{\alpha+1} - M_1M_\alpha) \end{aligned}$$

Here we use the shorthand notation $M_k = M_k(t) = \int_0^1 x^k \mu_t(x) dx$. If we now naively repeat the calculations we did in Section 2, we find the following:

$$\begin{aligned} \frac{d}{dt} M_1 &= s(M_2 - M_1) + u(M_{\alpha+1} - M_1 M_\alpha) \\ &= s(M_2 - M_1 - (M_{\alpha+1} - M_1 M_\alpha)) + (M_{\alpha+1} - M_1 M_\alpha)(s + u) \\ &= (M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha) \left[\left(\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \right) (s + u) - s \right] \end{aligned}$$

However, $\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha}$ is not relatedness if $\alpha \neq 1$. Moreover, if $\alpha \neq 1$, then also s and $s + u$ do not describe average costs and benefits for all densities μ_t . Therefore we first compute the correct average costs and benefits, which will be frequency dependent.

4.1 Costs and benefits

Costs are still constant; whatever the density μ_t is, being a cooperator implies giving up s in individual growth rate. The part of the benefits that is the result of the reduced individual growth rate is also constant. Benefits as a result of increased group reproduction, however, do depend on the group a cooperator is in. Average benefits therefore should be computed as a weighted average over all the types of groups a cooperator can be in.

In the limit of $n \rightarrow \infty$, the aggregate marginal change from being a cooperator rather than a defector is $\frac{d u x^\alpha}{d x} = u \alpha x^{\alpha-1}$ (note that for $\alpha = 1$ this equals u , which is independent of x). At a given density μ_t , that implies that the average aggregate benefit dispensed by being a cooperator instead of a defector through increased group growth rate equals:

$$\frac{\int_0^1 x u \alpha x^{\alpha-1} \mu_t(x) dx}{\int_0^1 x \mu_t(x) dx} = u \alpha \frac{M_\alpha}{M_1}$$

For $\alpha = 1$ this equals u , as we would expect.

4.2 Inclusive fitness with the standard definition of r

If we combine this with the standard r , then Hamilton's rule would have to be:

$$\begin{aligned}\frac{d}{dt}M_1 > 0 &\Leftrightarrow rb > c \\ \frac{d}{dt}M_1 > 0 &\Leftrightarrow \frac{M_2 - (M_1)^2}{M_1 - (M_1)^2} \left(s + u\alpha \frac{M_\alpha}{M_1} \right) > s\end{aligned}$$

However, above we have derived that the correct condition is:

$$\frac{d}{dt}M_1 > 0 \Leftrightarrow \left(\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \right) (s + u) > s$$

These conditions are the same if $\alpha = 1$, but in general they are not.

4.3 Inclusive fitness if r is allowed to vary

If we leave the definition of r open, and choose it such that $\frac{d}{dt}M_1 > 0$ coincides with $rb > c$, we should start with the equation that gives the correct prediction.

$$\frac{d}{dt}M_1 > 0 \Leftrightarrow \left(\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \right) (s + u) > s$$

This can be rewritten as

$$\frac{d}{dt}M_1 > 0 \Leftrightarrow \left(\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \frac{s + u}{s + u\alpha \frac{M_\alpha}{M_1}} \right) \left(s + u\alpha \frac{M_\alpha}{M_1} \right) > s$$

The costs are indeed s and the average benefits are indeed $s + u\alpha\frac{M_\alpha}{M_1}$. Hence the r that “makes Hamilton’s rule work” is:

$$r = \frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \frac{s + u}{s + u\alpha\frac{M_\alpha}{M_1}}$$

This does however depend on α , s and u .

4.4 Interpretation

In the previous subsection, we rewrote the condition for $\frac{d}{dt}M_1$ to be positive. The condition we arrived at is

$$\left(\frac{M_{\alpha+1} - M_1 M_\alpha}{M_1 - M_2 + M_{\alpha+1} - M_1 M_\alpha} \frac{s + u}{s + u\alpha\frac{M_\alpha}{M_1}} \right) \left(s + u\alpha\frac{M_\alpha}{M_1} \right) > s$$

It is a Procrustean stretch though to refer to this as “Hamilton’s rule”. Both r and the c/b ratio now depend on everything - both depend on α , s , u and the current density μ_t - and relatedness is therefore no longer a quantity that can be interpreted as a property of a population structure, as it varies with all parameters that describe the fitness effects. Therefore the “rule” itself changes along with every change in the case it is applied to, and r just serves as a remainder term to make up for whatever differences there are between what Hamilton’s rule would have to be with normal relatedness and the correct prediction.

5 Finding the trajectory with inclusive fitness

If we return to the case with $\alpha = 1$, then the time derivative of the first moment is positive if and only if $rb > c$, with r relatedness, and c and b costs and benefits;

$$\frac{d}{dt}M_1 = M_1(1 - M_1)(rb - c)$$

Even though c and b do not depend on the density μ_t , and therefore will be constant along the trajectory, for finding a solution of the PDE it is not enough to know relatedness at time $t = 0$. Knowing r at time $t = 0$ will tell us the direction of selection at time $t = 0$, but not at $t > 0$, because also r will change over time.

Here we compute the time derivative of r . This will also require that we compute $\frac{d}{dt} \int_0^1 x^2 \mu_t(x) dx = \frac{d}{dt} M_2$, the time derivative of the second moment. With $\frac{d}{dt} \int_0^1 x^2 \mu_t(x) dx = \int_0^1 x^2 \frac{d}{dt} \mu_t(x) dx$ we again exploit the PDE

$$\frac{d}{dt} \mu_t(x) = s \frac{d}{dx} [x(1-x) \mu_t(x)] + u \mu_t(x) \left[x - \int_0^1 y \mu_t(y) dy \right]$$

This is multiplied by x^2 and integrated from 0 to 1. The first term of the right hand side is then

$$\begin{aligned} s \int_0^1 x^2 \frac{d}{dx} [x(1-x) \mu_t(x)] dx &= s \left([x^3(1-x) \mu_t(x)]_0^1 - \int_0^1 [2x^2(1-x) \mu_t(x)] dx \right) = \\ &= -2s \int_0^1 [x^2(1-x) \mu_t(x)] dx = \\ &= 2s \left(\int_0^1 x^3 \mu_t(x) dx - \int_0^1 x^2 \mu_t(x) dx \right) = 2s(M_3 - M_2) \end{aligned}$$

The second term is:

$$\begin{aligned} u \int_0^1 x^2 \mu_t(x) \left[x - \int_0^1 y \mu_t(y) dy \right] dx &= u \left(\int_0^1 x^3 \mu_t(x) dx - \left(\int_0^1 x^2 \mu_t(x) dx \right) \left(\int_0^1 x \mu_t(x) dx \right) \right) \\ &= u (M_3 - M_2 M_1) \end{aligned}$$

Together that implies

$$\frac{d}{dt} \int_0^1 x^2 \mu_t(x) dx = \frac{d}{dt} M_2 = 2s (M_3 - M_2) + u (M_3 - M_2 M_1)$$

If we now take the time derivative of $r = \frac{M_2}{M_1} - \frac{M_1 - M_2}{1 - M_1}$, it is clear that this requires the time derivative of M_2 , which requires that we know M_3 . Similarly, the second derivative of r requires that we know M_4 , and so on. Therefore one needs to know all the moments of μ_0 to find the trajectory.

6 General baseline reproduction rates and steady states

If the baseline reproduction rates of groups and individuals are not the same, we include a parameter $w > 0$ that reflects the benchmark number of group events per individual event. The group reproduction rate as a function of the number of cooperators then becomes $w \left(1 + u \left(\frac{i}{n}\right)^\alpha\right)$. With this generalization, one can derive the PDE again, where the dynamics are determined by the value of $\rho = w \cdot \frac{u}{s}$.

$$\frac{d}{dt}\mu_t(x) = s \left(\frac{d}{dx} [x(1-x)\mu_t(x)] + \rho\mu_t(x) \left[x^\alpha - \int_0^1 y^\alpha \mu_t(y) dy \right] \right)$$

6.1 Steady states for $\alpha = 1$

A density μ is a steady state if

$$\frac{d}{dt}\mu(x) = 0 \text{ for all } x \in (0, 1)$$

This implies that for all $x \in (0, 1)$

$$\begin{aligned} \frac{d}{dx} [x(1-x)\mu(x)] + \rho\mu(x) \left[x - \int_0^1 y\mu(y) dy \right] &= 0 \\ \Downarrow & \\ \mu(x)(1-2x) + \mu'(x)(x-x^2) + \rho\mu(x)[x - M_1] &= 0 \end{aligned}$$

If we define $\mu(1) = \lim_{x \rightarrow 1} \mu(x)$, then also the following must hold

$$\begin{aligned}
-\mu(1) + \rho\mu(1)[1 - M_1] &= 0 \\
\mu(1)(\rho - 1 - \rho M_1) &= 0 \\
\mu(1) &= 0 \text{ or } M_1 = \frac{\rho - 1}{\rho}
\end{aligned}$$

If we take $M_1 = \frac{\rho - 1}{\rho}$ and fill that in in the PDE, we find that for all $x \in (0, 1)$:

$$\begin{aligned}
\mu(x)(1 - 2x) + \mu'(x)(x - x^2) + \rho\mu(x)\left[x - \frac{\rho - 1}{\rho}\right] &= 0 \Leftrightarrow \\
\mu(x)(1 + (\rho - 2)x - (\rho - 1)) + \mu'(x)(x - x^2) &= 0 \Leftrightarrow \\
\mu(x)(\rho - 2)(x - 1) + \mu'(x)(x - x^2) &= 0 \Leftrightarrow \\
\mu(x)(\rho - 2) - \mu'(x)x &= 0
\end{aligned}$$

Any $Cx^{\rho-2}$ is a solution of this differential equation, but we also need $\int_0^1 Cx^{\rho-2}dx = 1$

$$\int_0^1 Cx^{\rho-2}dx = 1 \Leftrightarrow C = \frac{1}{\int_0^1 x^{\rho-2}dx} = \frac{1}{\left[\frac{1}{\rho-1}x^{\rho-1}\right]_0^1} = \rho - 1$$

Therefore the following density is a steady state.

$$\mu(x) = (\rho - 1)x^{\rho-2}$$

The second moment and the r of the steady state are:

$$\begin{aligned}
M_2 &= (\rho - 1) \int_0^1 x^2 x^{\rho-2} dx = (\rho - 1) \int_0^1 x^\rho dx \\
&= (\rho - 1) \left[\frac{1}{\rho + 1} x^{\rho+1} \right]_0^1 = \frac{\rho - 1}{\rho + 1}
\end{aligned}$$

$$\begin{aligned}
r &= \frac{M_2}{M_1} - \frac{M_1 - M_2}{1 - M_1} = \frac{\frac{\rho-1}{\rho+1}}{\frac{\rho-1}{\rho}} - \frac{\frac{\rho-1}{\rho+1} - \frac{\rho-1}{\rho+1}}{1 - \frac{\rho-1}{\rho}} \\
&= \frac{\rho}{\rho + 1} - \frac{\rho - 1}{\rho + 1} = \frac{1}{\rho + 1}
\end{aligned}$$

The latter is quite intuitive, because according to Hamilton's rule - which points to the direction of selection correctly if $\alpha = 1$ - a steady state must satisfy $r = \frac{c}{b} = \frac{1}{1+\rho}$. Note however that this is a necessary, but not a sufficient condition for it to be a steady state; there are many densities with $r = \frac{1}{1+\rho}$, for which therefore $\frac{d}{dt}M_1 = 0$, but which are not steady states (for instance all those for which $M_1 \neq \frac{\rho-1}{\rho}$). Moreover, this steady state is stable.

6.2 Steady states if $\alpha > 0$

A density μ is a steady state if

$$\frac{d}{dt}\mu(x) = 0 \text{ for all } x \in (0, 1)$$

This implies that for all $x \in (0, 1)$

$$\begin{aligned}
\frac{d}{dx} [x(1-x)\mu(x)] + \rho\mu(x) \left[x^\alpha - \int_0^1 y^\alpha \mu(y) dy \right] &= 0 \\
\Updownarrow & \\
\mu(x)(1-2x) + \mu'(x)(x-x^2) + \rho\mu(x)[x - M_\alpha] &= 0
\end{aligned}$$

As before, $M_\alpha = M_\alpha(t) = \int_0^1 x^\alpha \mu_t(x) dx$.

If we define $\mu(1) = \lim_{x \rightarrow 1} \mu(x)$, then also the following must hold

$$\begin{aligned} -\mu(1) + \rho\mu(1)[1 - M_\alpha] &= 0 \\ \mu(1)(\rho - 1 - \rho M_\alpha) &= 0 \\ \mu(1) &= 0 \text{ or } M_\alpha = \frac{\rho - 1}{\rho} \end{aligned}$$

If we take $M_\alpha = \frac{\rho - 1}{\rho}$ and fill that in in the PDE, we find that for all $x \in (0, 1)$

$$\begin{aligned} \mu(x)(1 - 2x) + \mu'(x)(x - x^2) + \rho\mu(x) \left[x^\alpha - \frac{\rho - 1}{\rho} \right] &= 0 \Leftrightarrow \\ \mu(x)(1 - 2x + \rho x^\alpha - (\rho - 1)) + \mu'(x)(x - x^2) &= 0 \Leftrightarrow \\ \mu(x)(2 - 2x + \rho x^\alpha - \rho) + \mu'(x)(x - x^2) &= 0 \end{aligned}$$

This is a first order linear homogeneous equation which has the solution

$\mu(x) = C e^{\int_0^x p(y) dy}$, where, in this case, $p(y) = \frac{2 - 2y + \rho y^\alpha - \rho}{y(1 - y)}$. The constant C is found by setting $\int_0^1 \mu(x) dx = 1$, which results in the solution

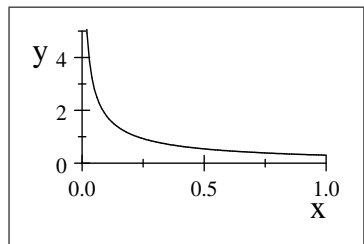
$$\mu(x) = \frac{e^{\int_0^x \frac{2 - 2y + \rho y^\alpha - \rho}{y(1 - y)} dy}}{\int_0^1 e^{\int_0^x \frac{2 - 2y + \rho y^\alpha - \rho}{y(1 - y)} dy} dx}$$

This does not always result in short expressions for all α , but some examples are:

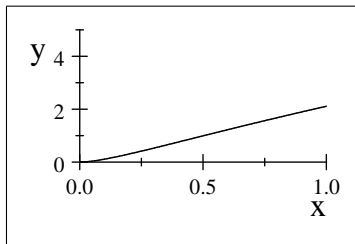
$$\begin{aligned} \alpha = 0.5 : \mu(x) &= \frac{\frac{x^{\rho-2}}{(1+\sqrt{x})^\rho}}{\int_0^1 \frac{x^{\rho-2}}{(1+\sqrt{x})^\rho} dx} \\ \text{From 6.1 } \alpha = 1 : \mu(x) &= (\rho - 1) x^{\rho-2} \\ \alpha = 2 : \mu(x) &= \frac{x^{\rho-2} e^{\rho x}}{\int_0^1 x^{\rho-2} e^{\rho x} dx} \\ \alpha = 3 : \mu(x) &= \frac{x^{\rho-2} e^{\frac{1}{2}\rho x(x+2)}}{\int_0^1 x^{\rho-2} e^{\frac{1}{2}\rho x(x+2)} dx} \end{aligned}$$

This allows us to draw a few examples of steady states.

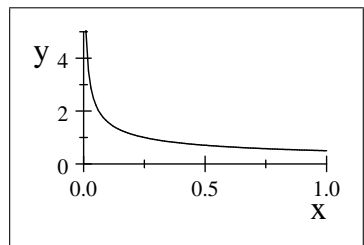
Figure 1 (next page). Stable steady states for different values of α and ρ . The shapes fit what we would intuitively expect; for higher α the stable steady densities are more curved, as the benefits of cooperation are more concentrated in the high fraction groups, while a higher ρ gives more cooperators in general.



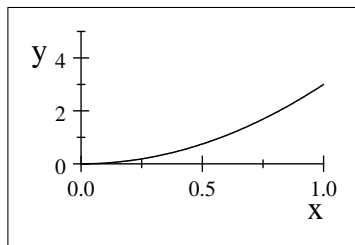
$\alpha = 0.5, \rho = 1.5$



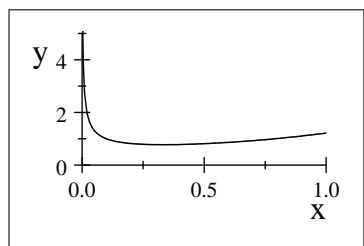
$\alpha = 0.5, \rho = 4$



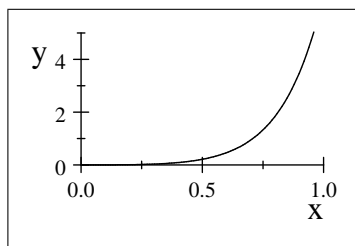
$\alpha = 1, \rho = 1.5$



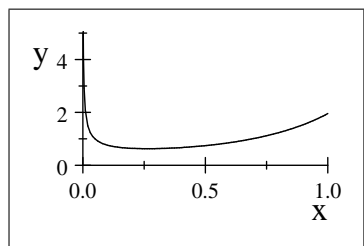
$\alpha = 1, \rho = 4$



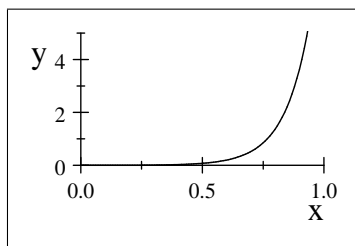
$\alpha = 2, \rho = 1.5$



$\alpha = 2, \rho = 4$



$\alpha = 3, \rho = 1.5$



$\alpha = 3, \rho = 4$

7 Related literature

7.1 Group selection, kin selection and inclusive fitness

There are many papers that comment on how group selection, kin selection and inclusive fitness relate. Our paper is about the relation between group selection and inclusive fitness. In the literature, however, the equivalence is referred to in different, seemingly interchangeable ways. Sometimes it is described as an equivalence of group selection and kin selection (Wade et al., 2010, Marshall, 2011). Other times the equivalence is described as a claim that any group selection models can always be reformulated in term of inclusive fitness. The second claim is explored in this paper. Even though kin selection and inclusive fitness are intimately related, these two “equivalences” are not the exact same claim. In general, the claims in the literature tend towards being somewhat verbal, rather than being exact and unambiguous mathematical claims. Therefore, in order to avoid all possible misunderstanding, and in order to delineate what our claims are, and what they are not about, we would like to include a description of how all three - group selection, kin selection and inclusive fitness - relate, or might relate, and we will very briefly touch upon some of the existing literature. Also we should stress here that this only describes theoretical possibilities, and as such it is not about whether or not there is empirical support that these possibilities describe actual processes that have shaped social evolution in one species or the other.

As a point of departure, we take Hamilton (1975). On page 336 of Hamilton (1975), he writes:

The usefulness of the ‘inclusive fitness’ approach to social behaviour (i.e. an approach using criteria like $(b_{AB}K - k) > 0$) is that it is more general than ‘group selection’, ‘kin selection’, or ‘reciprocal altruism’ approaches and so provides an overview even where regression coefficients and fitness effects are not easy to estimate or specify.

This can safely be taken to imply that both group selection and kin selection models are thought of as being included in the set of all models that can be described in terms of inclusive fitness. The paper by Hamilton (1975) is not about how group selection and kin selection relate, but contains some cues that suggest that neither one needs to be contained in the other. This is illustrated in Figure 2.

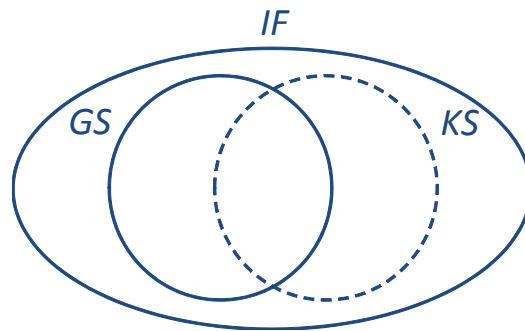


Figure 2. An illustration for Hamilton (1975)

To describe more in general how the three *could* relate, we draw a second figure (Figure 3). We have three sets of models; group selection models, kin selection models and models that can be analyzed with inclusive fitness. Any way in which those relate can be reduced to statements about subsets 1 to 7 in the figure.

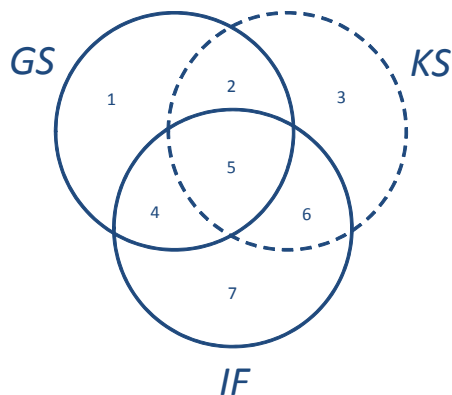


Figure 3. Possible ways in which group selection, kin selection and inclusive fitness relate are statements about these 7 subsets.

The quote from Hamilton (1975), depicted in Figure 1, basically suggests that subsets 1, 2 and 3 are empty, and that subset 7 is not. Our results disagree with the first part of that suggestion. We have shown that not all group selection models can be analyzed with inclusive fitness, and therefore that the union of sets 1 and 2 is *not* empty.

Before we will go over all the subsets, we should make two remarks. The first is about assortment and kin selection.

One of the key insights, or maybe *the* key insight in the theory of kin selection, is that the recipients of the benefits that cooperation brings might not be just anyone. Others that share a gene for altruism, or cooperation, might be overrepresented on the receiving end. For a selfish gene, making its bearer be an altruist might therefore be in its own genetic interest, if the recipient of that altruism is relatively likely to also be a carrier of that gene. Moreover, if a trait is truly altruistic - that is, if it involves an actual net cost; this is sometimes referred to as *strong* altruism - then the flip side of the same coin is that it will not be selected, unless there is some way in

which the average altruist faces a different set of recipients than the average non-altruist.

Here one can properly say, almost trivially, that assortment is more general than genetic assortment. If there is some mechanism that assorts altruistic behaviors, then those can be selected, even if there is no genetic assortment.¹ While Doebeli & Fletcher (2009) suggest to reserve the word kin selection for assortment that is genetic, Gardner, West & Wild (2011) include all sources of common descent when the word kin selection is used. Whichever choice is made is inconsequential for our findings, which concern inclusive fitness, not kin selection. For this overview we however choose to reserve the word kin selection for cases where assortment is genetic. If we do, then that implies that the union of 1 and 4 is non-empty - we allow for assortment in groups to be non-genetic.

What should be included in the set of group selection models is another question. Here the aim is not to determine the exact boundaries of that set. We do however want to exclude models that feature groups, but that are not group selection models, and in which the properties of individuals in the group are just the result of individual selection. One possible definition we could use to exclude those is the definition from Simon et al. (2013). That definition states that *“if a trait establishes itself in a model of two-level population dynamics when group-level events are present, and does not establish itself in the same model when they are absent, then the trait evolves by group selection.”* The model from the main text clearly satisfies that definition. There are also other models, in which group level events are not explicitly modeled. For those this definition is hard to check, as it is unclear how to take the group level events out. Yet they may very well qualify for being a group selection model. In the discussion below, we will argue that

¹A mechanism that picks phenotypes and makes them interact preferentially could induce some genetic assortment, but if there are multiple genes in the population that code for the same behaviour, the genetic assortment is less than the phenotypic assortment, and with such a mechanism the latter is relevant for selection.

all subsets are non-empty. For the purpose of this discussion, therefore, it is enough to point to models that surely should be included according to any definition of group selection. Which other models should also be allotted to those sets is also interesting, but not the aim of this exercise.

Subset 5. We would like to begin the overview with subset 5. This set contains models that fall both under group and kin selection, and moreover can be analyzed with inclusive fitness. The current consensus is that *all* group selection models can be analyzed with inclusive fitness. Given that this equivalence is regularly also referred to as the equivalence of group and kin selection, it is completely uncontroversial to say that this set is not empty. An actual theorem that shows that this is indeed not empty is Theorem 5 in Van Veelen (2011a). If relatedness is exogenously fixed, and the interactions have fitness effects that imply “generalized equal gains from switching” (Nowak & Sigmund, 1990, Van Veelen 2009, 2011a), then the theorem shows that inclusive fitness actually determines the entire dynamics. If assortment is furthermore genetic, that puts us into subset 5.

Subset 4. If assortment is not genetic, but the fitness effects imply generalized equal gains from switching, then the same theorem applies, but the model finds itself in subset 4. This subset therefore is not empty either. Again, whether or not this possibility is relevant depends on whether or not there are interesting examples of non-genetic assortment in groups.

Subsets 1 and 2. The results in this paper imply that the union of sets 1 and 2 is not empty; there are group selection models that can not be recast in terms of inclusive fitness. If we allow for assortment of phenotypes as well as genotypes, that means that both 1 and 2 individually are not empty either.

Subsets 6 and 7. Subset 6 is generally accepted not to be empty. Lion, Jansen & Day (2011) point out that with local interaction models - which are not group selection models - inclusive fitness can work, which implies that the union of 6 and 7 is not empty. If the local interaction dynamics

reflect physical reproduction, then the model is in subset 6, if it reflects social learning or imitation, then the model is in subset 7.

Subset 3. Whether subset 3 is non-empty is perhaps more of an open question. If replication in local interaction structures is physical, then Part A of the SI of Nowak, Tarnita & Wilson (2010) - which is about symmetric interactions in structured populations, and *not* about eusociality - can be interpreted as to imply that subset 3 is not empty. Their setup typically also allows for an interpretation in terms of social learning, so that claim can also be interpreted as a statement that there are models that are in none of the three sets; neither group selection, nor kin selection, nor allowing for an inclusive fitness analysis.

7.2 Empirics

There is no consensus about whether or not group selection is helpful, or needed, for explaining any observed phenomenon in the living world, and especially in humans. Positions range from Sober & Wilson (1998) and Wilson & Wilson (2007), who claim that group selection is an essential ingredient in explaining many cooperative behaviours, including in humans, to Dawkins (1989) and Pinker (2012), who claim it contributes nothing to our understanding of them. Although there are obviously links with theory, this is mainly an empirical question. Answering it would involve estimates of parameters that feature in theoretical group selection models in order to see if the circumstances in the real world are comparable, or even anywhere close to those that are required to make the theoretical model work. Also the predictions that group selection models make, for instance about behaviour, should be tested with data, and this should be compared to how well these data are in line with alternative explanations. This paper is not about that empirical question. Our paper is only an investigation whether or not any group selection model can always be reformulated in terms of inclusive fitness. We claim that this is true in some settings, but not in others.

References

- [1] Bergstrom T (2003) The algebra of assortative encounters and the evolution of cooperation. *Int Game Theory Rev* 5:211-228.
- [2] Bowles S, Gintis H (2011) *A cooperative species—human reciprocity and its evolution* (Princeton University Press, Princeton).
- [3] Dawkins R (1989) *The Extended Phenotype*. (Oxford University Press, Oxford).
- [4] Fletcher JA, Doebeli M (2009) A simple and general explanation for the evolution of altruism *Proc R Soc B* 276:13-19.
- [5] Durrett R (2008) *Probability Models for DNA Sequence Evolution*. 2nd edition. (Springer, New York)
- [6] Eshel I, Cavalli-Sforza LL (1982) Assortment of encounters and evolution of cooperativeness. *Proc Natl Acad Sci USA* 79:1331–1335.
- [7] Gardner A, West SA, Wild G (2011) The genetical theory of kin selection. *J Evol Biol* 24:1020-1043.
- [8] Grafen A (1985) A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2:28-90.
- [9] Hamilton WD (1975) Innate social aptitudes of man: Approach from evolutionary genetics. Pages 133-155 in Fox, R., ed. *Biosocial Anthropology* (Wiley, New York).
- [10] Karlin S, Taylor HM (1981) *A Second Course in Stochastic Processes*. (Academic Press, New York - San Francisco - London).
- [11] Lion S, Jansen VAA, Day T (2011) Evolution in structured populations: beyond the kin versus group debate. *Trends Ecol Evol* 26:193-201.

- [12] Luo SA (2014) A unifying framework reveals key properties of multilevel selection. *J Theor Biol* 341:41–52
- [13] Marshall JAR (2011) Group selection and kin selection: formally equivalent approaches. *Trends Ecol Evol* 26:325-332.
- [14] Nowak MA, Sigmund K (1990) The evolution of stochastic strategies in the prisoner's dilemma." *Acta Appl Math* 20:247-265.
- [15] Nowak MA, Tarnita CE, Wilson EO (2010) The evolution of eusociality, *Nature* 466:1057–1062.
- [16] Orlove MJ, Wood CL (1978) Coefficients of relationship and coefficients of relatedness in kin selection: a covariance form for the RHO formula. *J Theor Biol* 73:679–686.
- [17] Pinker S (2012) The false allure of group selection, *Edge*, June 21, 2012, <http://edge.org/conversation/the-false-allure-of-group-selection>
- [18] Queller DC (1992) Quantitative genetics, inclusive fitness, and group selection. *Am Nat* 139:540-558.
- [19] Simon B, Fletcher JA, Doebeli M (2013) Towards a general theory of group selection, *Evolution* 67:1561–1572.
- [20] Sober, E, Wilson DS (1998) *Unto Others; the evolution and psychology of unselfish behavior*. (Harvard University Press, Cambridge, MA).
- [21] van Veelen M (2009) Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *J Theor Biol* 259:589-600.
- [22] van Veelen M (2011a) A rule is not a rule if it changes from case to case (a reply to Marschall's comment). *J Theor Biol* 270:189-195.
- [23] van Veelen, M (2011b). The replicator dynamics with n player games and population structure. *J Theor Biol* 276:78-85.

- [24] Wade MJ, Wilson DS, Goodnight C, Taylor D, Bar-Yam Y, de Aguiar MAM, Stacey B, Werfel J, Hoelzer GA, Brodie III ED, Fields P, Breden F, Linksvayer TA, Fletcher JA, Richerson PJ, Bever JD, Van Dyken JD, Zee P (2010) Multilevel and kin selection in a connected world. *Nature* 463:E8-E9.
- [25] Wilson DS, Wilson EO (2007) Rethinking the theoretical foundations of socio-biology. *Q Rev Biol* 82:327-348.