



## Letter to Editor

**A rule is not a rule if it changes from case to case (a reply to Marshall's comment)**

## ARTICLE INFO

*Keywords:*

Hamilton's rule  
 Queller's rule  
 Inclusive fitness  
 Group selection  
 Generalized equal gains from switching

## ABSTRACT

This is a reply to "Queller's rule ok: Comment on van Veelen 'when inclusive fitness is right and when it can be wrong' " by James Marshall in the Journal of Theoretical Biology, in this issue.

In order to circumvent the disagreement about the Price equation and focus on the issue of the predictive power of inclusive fitness for group selection models, I derive Queller's and Marshall's rule without the Price equation. Both rules however need a translation step in order to be able to link them to the group selection model in van Veelen (2009). Queller's rule applies to games with 2 players and 2 strategies, and is general. Marshall's rule on the other hand applies only to a small subset of 3-player games. His rule is correct, but for other, similarly small subsets we would get other rules. This implies that if we want a rule that applies to *all* symmetric games with 3 players and 2 strategies, it will have to use a vector of dimension 2 that represents population structure. More in general: for group selection models with groups of size  $n$ , a correct and general prediction will need to use a vector of dimension  $n - 1$  that represents population structure.

© 2010 Elsevier Ltd. All rights reserved.

**1. Introduction**

There are two points on which Marshall (this issue) and I differ. The first is that I think that however well the Price equation can serve as a carrier of good intuition, it just fails to provide a formally sound method for modelling as well as for statistical inference. This is described in detail in van Veelen (2005). Marshall disagrees with that. The second is that I claim that a prediction for a general group selection model cannot be (1) general, (2) always correct and (3) based on less than  $n - 1$  numbers that characterize population structure, where  $n$  is the group size (van Veelen, 2009). Marshall also disagrees with that. These two disagreements together constitute a bit of Gordian knot. When Marshall claims that my Counterexample II (a 3-player game meant to show that 1 parameter that reflects population structure is not enough) is not a real counterexample, he uses a Price equation approach that he borrows from Queller (1985, 1992) in order to arrive at a rule with 1 parameter only. If one disagrees with derivations that use the Price equation, then it is hard to be convinced by a claim that is arrived at with exactly that method. But simply dismissing his claim on methodological grounds I think would not do justice to the point he raises. I will therefore focus on the second issue and try to find a way around the point concerning the Price equation.

What is a bit unfortunate is that Marshall, in his discussion of my counterexample, remains strictly within Queller's framework and his use of variables. I use variables such as  $f_i$  (the frequency of groups with  $i$  cooperators in it),  $p$  (the frequency of the gene in the overall population) and  $r$ , while Marshall uses covariances for which it is not clear how these should relate to my variables. At first sight they should even make us wonder if they relate at all. Note that my model is completely deterministic, which implies that whatever variables we think of, all their variances and covariances are 0. Yet Marshall's formulas contain covariances that are certainly not all meant to be 0. Below we will see that a link between the two can in fact be established, but we will need to find a way to properly translate

one into the other. In Sections 3 and 4 we will therefore start with a very simple assumption about the dynamics (payoff monotonicity) and show how both Queller's and Marshall's rule follow from this assumption, without the use of the Price equation. Both rules now use my variables, and the appendix shows how the rules stated in Queller's variables can be translated to mine and back.

Once we know how to translate results stated in terms of Queller's (1985, 1992) into results that are derived without reference to the Price equation, we will find that Marshall's claim concerning my Counterexample II game is, with a bit of rephrasing, correct. Indeed, for his choice of a subset of games he derives a rule that has only a scalar in it that represents the population structure. But his conclusion misses the point that I should have made more clearly. The point is that the correct rule is different for different subsets of games. In other words, if we change the payoffs of the game, then not only do we have to change the numbers we fill in the rule (which reflect those payoffs), but we also have to change the rule itself, if we want it to give the correct prediction.

This disagreement is an instance of a wider debate on the value of inclusive fitness (see for example Traulsen and Nowak, 2006; Lehmann et al., 2007; Killingback et al., 2006; Grafen, 2007; Wild et al., 2009; Wade et al., 2009; Nowak et al., 2010). There is not much convergence so far, but I think that this exchange can help improve the communication between the different sides in the debate. Especially the translation exercise should serve as an example that hopefully will facilitate the exchange of ideas.

**2. What is population structure here, and what is inclusive fitness?**

Although there are perhaps different opinions on what the basic idea of inclusive fitness is, I think that it is most useful to think of it as a way to separate two things: on the one hand there is a

population or interaction structure, and on the other hand there are the fitness effects on self and others.

In my simple group selection model, groups are all of size  $n$ . For simplicity we assume an infinite population, which implies that the number of groups is also infinite. Individuals are either cooperators or defectors,<sup>1</sup> which implies that the type of group can uniquely be described by the number of cooperators in it. A particular composition of the population therefore is given by  $n+1$  numbers  $f_0, \dots, f_n$ , which represent the frequency of groups with  $0, \dots, n$  cooperators in it. There are of course restrictions on these numbers; they are frequencies, so they must lie between 0 and 1 and they have to add up to 1. In formulas:  $0 \leq f_i \leq 1$  for all  $i$ , and  $\sum_{i=0}^n f_i = 1$ . The gene frequency  $p$  in the overall population is given by  $p = (1/n) \sum_{i=0}^n i f_i$ . If indeed  $0 \leq f_i \leq 1$  for all  $i$ , then the restriction  $\sum_{i=0}^n f_i = 1$  implies that knowing only  $n$  of those  $n+1$  numbers is enough, because the last one follows from the restriction. In case the frequency  $p$  is also given, knowing only  $n-1$  of those  $n+1$  numbers is enough for the same reason.

On the other hand we have the fitness effects, which in my paper are reflected by game payoffs. In the paper I allow those to be frequency dependent, but for expositional clarity it can help to assume that they are just fixed numbers. Here I denote these numbers with  $\pi_{C,i}$  and  $\pi_{D,i}$ , which are the payoffs to a cooperator and a defector, respectively, if they find themselves in a group which has, in total,  $i$  cooperators in it. Of course  $\pi_{C,0}$  and  $\pi_{D,n}$  cannot exist—there are no cooperators in groups with 0 cooperators and no defectors in groups with  $n$  cooperators—so  $\pi_{C,i}$  is defined for  $i=1, \dots, n$  and  $\pi_{D,i}$  is defined for  $i=0, \dots, n-1$ . With group size  $n$  this makes  $2n$  numbers.

There is one simple and natural assumption that we make about the dynamics. We assume that if cooperators on average have a higher payoff than defectors, then their frequency increases, and if they on average have a lower payoff than defectors, then their frequency decreases. This assumption is known as payoff monotonicity, and it is probably the most general and modest assumption one can make. For instance Vega-Redondo (2003, p. 377) writes that “payoff monotonicity is usually interpreted as reflecting the minimal criterion of evolutionary consistency that any reasonable model of evolution, social or otherwise, should satisfy”. See also Hofbauer and Sigmund (1998, Section 8.2, pp. 88–89).

In order to determine the direction of selection we therefore only look at the average payoff to a cooperator and the average payoff to a defector, compare them, and require that the dynamics respect the sign of the difference. This defines payoff monotonic dynamics and because we will use this property at a few more instances, it is stated below in a formula.

PAYOFF MONOTONICITY:

$$\dot{p} > 0 \text{ if and only if } \frac{\sum_{i=1}^n i \cdot f_i \cdot \pi_{C,i}}{np} > \frac{\sum_{i=0}^{n-1} (n-i) \cdot f_i \cdot \pi_{D,i}}{n(1-p)}$$

This is satisfied for instance by the replicator dynamics as a special case with  $n=2$  and random matching. Random matching here means that, for a given frequency  $p$  in the overall population,  $f_0=(1-p)^2$ ,  $f_1=2p(1-p)$  and  $f_2=p^2$ , and a simple calculation returns that the requirement would reduce to “ $\dot{p} > 0$  if and only if  $p\pi_{C,2} + (1-p) \cdot \pi_{C,1} > p\pi_{D,1} + (1-p)\pi_{D,0}$ ” which is obviously implied by the replicator dynamics.<sup>2</sup>

<sup>1</sup> In general games with two actions it is not necessarily clear or even well defined what the cooperative action is and what the defective action is. In the examples here it is clear, so here I will stick to these terms, but we can read it as an action C and an action D, where C and D are just labels of two different behaviours or strategies.

<sup>2</sup> Payoff monotonicity does not restrict the direction of the dynamics for  $p=0$  and  $p=1$ , since the average payoff of cooperators resp. defectors cannot be computed there. A natural extension would be to require that  $\dot{p} = 0$  for  $p=0$  and  $p=1$ .

It is however interesting to note that our requirement is more general in two ways. It allows for groups of any size as well as for non-random matching. Frequencies  $f_i = \binom{n}{i} p^i (1-p)^{n-i}$  for  $n=0, \dots, n$  constitute an example of groups of any size with random matching. An example of non-random matching with groups of 2 would be  $f_0=(1-p)$ ,  $f_1=0$  and  $f_2=p$ , which would represent clonal interaction. For groups of any size that would generalize to  $f_0=(1-p)$ ,  $f_1, \dots, f_{n-1}=0$  and  $f_n=p$ .

A final thing to note here is that if we add a constant  $a$  to all payoffs  $\pi_{C,i}$  and  $\pi_{D,i}$ , then both sides in the condition for payoff monotonicity increase by  $a$ . This implies that if  $\dot{p} > 0$  for payoffs  $\pi_{C,i}$  and  $\pi_{D,i}$ , then also  $\dot{p} > 0$  for payoffs  $\tilde{\pi}_{C,i} = \pi_{C,i} + a$  and  $\tilde{\pi}_{D,i} = \pi_{D,i} + a$ . We can therefore normalize the payoffs (for instance such that  $\pi_{D,0} = 0$  or  $\pi_{D,0} = 1$ ), which leaves us with  $2n-1$  parameters to choose.

Even though most of us will have an intuition for what inclusive fitness should be, it is not that easy to find, or even make a proper definition. It is however of crucial importance to have at least a starting definition, because if we do not delineate what is and what is not a prediction based on inclusive fitness, then we surely cannot answer the question whether inclusive fitness (always) gives the correct prediction. It is therefore also of importance to think, if inclusive fitness is to be meaningful as a term, what a prediction would look like if it were *not* based on inclusive fitness.

I suggest we start by looking at Hamilton’s rule, surely an example of a rule based on inclusive fitness. Hamilton’s rule (altruistic behaviour is selected if  $rb-c > 0$ ) can be seen as a prediction based on the comparison of the inclusive fitness of the altruistic behaviour ( $rb-c$ ) and the inclusive fitness of not behaving altruistically ( $r \cdot 0 - 0 = 0$ ). There are a few perhaps all too obvious observations I would like to make. The first is that the  $r$  only reflects population structure, and that  $b$  and  $c$  only reflect fitness effects. The second is that the condition could also be written as  $r-c/b > 0$  or  $r > c/b$ , which reduces the number of variables reflecting fitness effects to 1 (only  $c/b$ ). Also note that this condition is not frequency dependent; the frequency of the gene does not feature in the prediction (it actually did in Hamilton’s, 1964a,b first result, that states that *mean* inclusive fitness is maximized in the population, but not in his 1964b second claim, which is that individuals maximize their inclusive fitness. See Van Veelen (2007) for the link between the two).

Suppose the set of models we have implies that we can choose from a set  $S$  of population structures on the one hand, and from a set  $\Pi$  of fitness effects on the other. If the condition is to look like Hamilton’s rule, then the condition can be constructed in two steps; first we make a number that reflects the population structure ( $r : S \rightarrow \mathbb{R}$ ) and then we make a number that reflects fitness effects ( $d : \Pi \rightarrow \mathbb{R}$ ). Then we compare these two numbers in order to determine the direction of selection.

Hamilton’s rule : The trait (behaviour) is selected if  $r(s) > d(\pi)$

I would like to contrast that to what any rule would look like. In the general form for a rule below, we have a function  $g : S \times \Pi \rightarrow \mathbb{R}$ .

Any rule : The trait (behaviour) is selected if  $g(s, \pi) > 0$

Two things are important to notice here. The first is that these rules will have to hold for any combination of a  $s \in S$  and a  $\pi \in \Pi$ . The second is that if  $S$  and  $\Pi$  are both of dimension  $n > 2$ , then not all conditions of the form  $g(s, \pi) > 0$  can be rewritten as  $r(s) > d(\pi)$ . Just to give an extremely simple example, suppose that  $S = \Pi = \mathbb{R}^3$ . Then the condition  $s_1 \pi_1 + s_2 \pi_2 + s_3 \pi_3 > 0$  cannot be rewritten in a form where population structure and fitness effects are separated by an inequality sign.

Now I think that any meaningful definition of inclusive fitness would have to be a restriction on the form of the condition in this prediction. If inclusive fitness is not a restriction, then it is

meaningless; it would just become a synonym for a condition, or, if one adds the requirement that it must give the correct prediction, for a condition that gives the correct prediction. So restrictions on the functional form are a necessity. There are a few restrictions one can think of. The first is that inclusive fitness must separate population structure from fitness effects in two 1-dimensional numbers, as Hamilton's rule does.

Restriction 1 : The condition must be equivalent to  $r(s) > d(\pi)$ .

where  $r : S \rightarrow \mathbb{R}$  and  $d : \Pi \rightarrow \mathbb{R}$ . This means that if the function  $g$  in the condition  $g(s, \pi) > 0$  looks like  $g(s, \pi) = r(s) - d(\pi)$ , then Restriction 1 is satisfied.

It is also important to note that if the correct rule cannot be rephrased as in Restriction 1 on the whole of  $S \times \Pi$ , it could be that there is a subset of  $S' \times \Pi' \subset S \times \Pi$  for which one can rephrase the correct rule as such. One such restriction is that the games must satisfy generalized equal gains from switching (Theorem 1 in van Veelen, 2009), but there is a range of possibilities, and we will see an alternative subset of games below.

Another restriction one could think of is frequency independence.

Restriction 2 : The condition must be frequency independent

In Section 3 we will see that Queller's rule satisfies neither of these two restrictions.

Of course I am open to all suggestions for restrictions other than those that could define what inclusive fitness is. But it is worth emphasizing that only restrictions on the functional form of the condition can give the concept of inclusive fitness meaning; inclusive fitness can only be a meaningful concept if there could also be rules that are *not* based on it.

### 3. Queller's rule

We begin by comparing the payoff matrices implied by Queller (1985, p. 366) and Counterexample I in van Veelen (2009, pp. 594–595).

		Not altruistic	Altruistic
Queller : Not altruistic		$W_0$	$W_0 + B$
	Altruistic	$W_0 - C$	$W_0 + B - C + D$
		D	C
van Veelen : D	P	T	
	C	S	R

Because we assume the dynamics to be payoff monotonic, we can subtract  $W_0$  from all payoff entries in Queller without consequences. This implies that the matrix—with lower case letters replacing capitals—becomes

		Not altruistic	Altruistic
Queller : Not altruistic		0	$b$
	Altruistic	$-c$	$b - c + d$

If we assume payoff monotonicity, which assumes that selection favours the trait with the higher average payoff, then cooperation is favoured if:

$$\frac{f_2 \cdot \mathbf{R} \cdot 2 + f_1 \cdot \mathbf{S} \cdot 1}{2p} > \frac{f_1 \cdot \mathbf{T} \cdot 1 + f_0 \cdot \mathbf{P} \cdot 2}{2(1-p)} \tag{C1}$$

This is Eq. (3) in van Veelen (2009, p. 594). We can however replace the payoffs by their Queller counterparts, in order to arrive at

$$\frac{f_2 \cdot (b - c + d) \cdot 2 + f_1 \cdot -c \cdot 1}{2p} > \frac{f_1 \cdot b \cdot 1 + f_0 \cdot 0 \cdot 2}{2(1-p)}$$

Singling out the  $c$ , the  $b$  and the  $d$ , this is rewritten as

$$-\frac{2f_2 + f_1}{2p}c + \left(\frac{2f_2}{2p} - \frac{f_1}{2(1-p)}\right)b + \frac{2f_2}{2p}d > 0.$$

Now we use  $p = (2f_2 + f_1)/2$ ,  $r = 2f_2/2p - f_1/2(1-p)$  and  $f_2/p = \mathbb{P}(T|T) = r + (1-r)p$  (see van Veelen, 2009 and Appendix A) to arrive at

QUELLER'S CONDITION

$$-c + rb + (r + (1-r)p)d > 0$$

In Appendix A we can see that this is exactly what we get if we translate Queller's rule, stated in terms of covariances, to the deterministic dynamics, according to a hypothetical chance experiment. It is also duly noted that with equal gains from switching ( $d=0$  or  $\mathbf{R} - \mathbf{S} = \mathbf{T} - \mathbf{P}$ ) Hamilton's rule reappears.

Now what do we learn from this derivation without the use of the Price equation? First of all I would say that it is worth to note that Queller's rule follows from an assumption; if payoff monotonicity is not satisfied, Queller's rule does not follow. Or, more precisely, in the case of a game with 2 players and 2 strategies, payoff monotonicity—reflected by C1 for this special case—and Queller's rule are actually equivalent, because we could also do the derivation in reverse.

Given that they are equivalent, I think is a matter of taste which of the two one would like to use, where Queller's condition perhaps looks a bit shorter than C1. However, if we are going to increase the size of the group, then, as Queller (1985, p. 367) recognised, the condition will have to be extended. While the generalization of C1 still looks quite manageable (it is just payoff monotonicity;  $(1/np) \sum_{i=1}^n i \cdot f_i \cdot \pi_{C,i} > (1/n(1-p)) \sum_{i=0}^{n-1} (n-i) \cdot f_i \cdot \pi_{D,i}$ ), my expectation is that without restrictions on the admissible payoffs or possible group compositions, there is no straightforward generalization of Queller's condition that looks anywhere attractive. In the same way as  $d$  is added to replace Hamilton's rule by the more general Queller's rule, we would then go on adding ever more variables, the coefficients of which in the ever more general rule get ever more complicated. It remains a matter of taste, but the general version of Queller's condition will at some point certainly not be shorter than the general version of C1.

It is also worthwhile to focus on the degrees of freedom in the choice of a population structure and the choice of a game, here and further below. With parameters  $b$ ,  $c$  and  $d$ , Queller has introduced maximum generality concerning the payoffs in 2-player games with 2 actions, if we only consider payoff monotonicity as a property of dynamical systems. If the average payoff of the one trait is larger than the average payoff of the other for the matrix  $\begin{bmatrix} 0 & b \\ -c & b-c+d \end{bmatrix}$ , then the same holds for  $\begin{bmatrix} a & a+b \\ a-c & a+b-c+d \end{bmatrix}$ . So his rule accommodates all possible payoff matrices, if we normalize the payoffs, for instance such that the payoff in the upper left corner of the payoff matrix is 0.

Finally it is important to realise that if we consider a given frequency  $p$ , then there is a 1-dimensional threshold in the population structure. For a given game (fixed  $b, c$  and  $d$ ), and a given frequency  $p$ , the condition  $-c + rb + (r + (1-r)p)d > 0$  separates all possible values of  $r$  in two sets: on the one hand the  $r$ 's for which the condition holds, and for which the trait or behaviour is selected, and on the other hand the  $r$ 's for which the condition does not hold, and for which the trait is not selected. This is not surprising, because with  $n=2$ , there is only one degree of freedom in the choice of  $f_0, f_1$  and  $f_2$ . After all, they must satisfy  $f_0 + f_1 + f_2 = 1$  and  $f_2 + \frac{1}{2}f_1 = p$ . Choosing  $f_1$  and  $f_2$  is equivalent to choosing  $p$  and  $r$ —with  $r = 2f_2/2p - f_1/2(1-p)$ —so with  $p$  given, only  $r$  can be chosen. If  $p$  is not given, then Restriction 1 cannot be satisfied, and the threshold will have to be two dimensional.

4. Marshall's rule

The game for which Marshall derives his rule is reproduced below (the assumption that  $b=0$  is already incorporated here). As he notes, my Counterexample II corresponds to  $c=1$  and  $d=2$ . Note that the number of parameters for this game is 2, while it is 5 for the general game, if we normalize the payoffs such that  $\pi_{D,0} = 0$ .

		0 other cooperators	1 other cooperators	2 other cooperators
Marshall :	Defect	0	0	0
	Cooperate	-c	-c	-c+d
		0 other cooperators	1 other cooperators	2 other cooperators
General :	Defect	$\pi_{D,0}$	$\pi_{D,1}$	$\pi_{D,2}$
	Cooperate	$\pi_{C,1}$	$\pi_{C,2}$	$\pi_{C,3}$

If we again assume payoff monotonicity, and fill in Marshall's payoffs, then we find that  $\dot{p} > 0$  if and only if

$$\frac{\sum_{i=1}^n i \cdot f_i \cdot \pi_{C,i}}{np} > \frac{\sum_{i=0}^{n-1} (n-i) \cdot f_i \cdot \pi_{D,i}}{n(1-p)} \Leftrightarrow$$

$$\frac{\sum_{i=1}^2 i \cdot f_i \cdot (-c) + 3 \cdot f_3 \cdot (-c+d)}{np} > 0 \Leftrightarrow$$

$$\frac{-cnp + 3df_3}{np} > 0 \Leftrightarrow$$

$$\frac{df_3}{p} > \frac{cp}{p}$$

This reduces to

MARSHALL'S CONDITION

$$\frac{f_3}{p} > \frac{c}{d}$$

Appendix B shows how this is exactly what we get if we translate the rule from Marshall's comment, stated in terms of covariances, to the deterministic dynamics, according to a hypothetical chance experiment. This implies that Marshall's rule, with an interpretation of variance and covariance as belonging to a hypothetical chance experiment, is indeed the appropriate rule for a subset of 3-player games defined with the two parameters  $c$  and  $d$ . For this subset of 3-player games, indeed the condition is based on a scalar that follows from the population structure— $f_i/p$ . Now the question remains whether or not this implies that my example is not an actual counterexample. This will be discussed in the following section.

5. Claims and counterexamples

The point I tried to make in my paper is the following. Suppose we take a fixed group size  $n$ . Then there is on the one hand a set of population structures and on the other hand a set of games. The set of population structures has dimension  $n - 1$  if we do, and dimension  $n$  if we do not fix frequency  $p$ . The set of games has dimension  $2n - 1$  if we normalize. If we assume payoff monotonicity, which is only natural and holds for instance for the replicator dynamics, then we get a condition for selection to favour trait C over trait D that has the following shape:

Trait C is selected if  $g(s, \pi) > 0$ ,

where  $s \in S$  is any population structure and  $\pi \in \Pi$  any game, and where  $g(s, \pi)$  reflects the assumption of payoff monotonicity;

$$g(s, \pi) = \frac{\sum_{i=1}^n i \cdot f_i \cdot \pi_{C,i}}{np} - \frac{\sum_{i=0}^{n-1} (n-i) \cdot f_i \cdot \pi_{D,i}}{n(1-p)}$$

My claim was that there is no rule that is (1) general, (2) always correct, and (3) based on less than  $n - 1$  numbers that characterize population structure, where  $n$  is the group size (van Veelen, 2009). Maybe I can be more precise here. By general, I mean that the rule applies to all population structures  $s \in S$  and all games  $\pi \in \Pi$ , which in this case are sets of dimensions  $n$  and  $2n - 1$ . By correct I mean that it respects payoff monotonicity. By the third requirement I mean that the rule uses a vector of dimension lower than  $n - 1$  to represent population structure. This certainly implies that if  $n > 2$ , this rule cannot depend on population structure through a single number  $r$ . In other words, it is not possible that the condition is equivalent to

Trait C is selected if  $r(s) > d(\pi)$

with  $r : S \rightarrow \mathbb{R}$  and  $d : \Pi \rightarrow \mathbb{R}$ .

If we lift any of these three requirements, then there are rules that satisfy the other two. If we restrict the generality, and reduce for instance the set of games, then we can arrive at rules that are correct, and feature population structure reduced to a scalar. One such example is my Theorem 1. This theorem restricts the set of games to those that satisfy generalized equal gains from switching (which economists refer to as linear public goods games). This restriction implies that the set of games is characterized by only two parameters ( $b$  and  $c$ ) instead of  $2n - 1$ . For this subset of games, the theorem shows the following rule to be correct for any group size  $n$ :

Trait C is selected if  $r > \frac{c-b}{(n-1)b}$

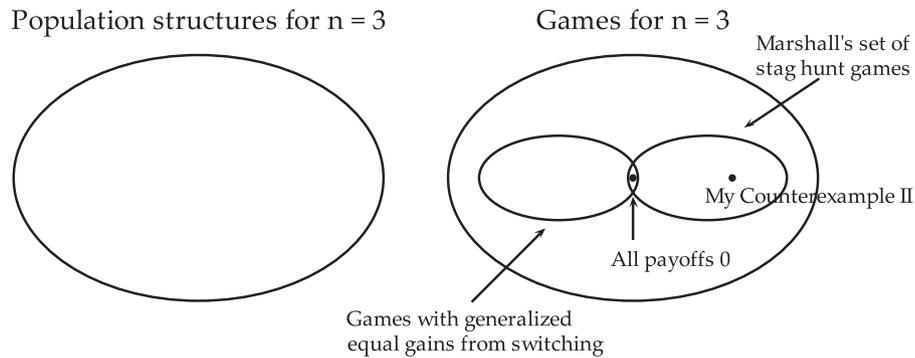
Here  $r = \mathbb{P}(T|T) - \mathbb{P}(T|N)$  as defined in the paper, which is indeed a function of population structure only, because it only depends on  $f_0, \dots, f_n$ .

Another restriction on the set of games is the one chosen by Marshall in order to include my Counterexample II. He restricts attention to a 2-dimensional subset of the 5-dimensional set of games with 3 players and 2 strategies, namely the set of stag-hunt games that is characterized by the parameters  $c$  and  $d$ . For that subset, we have seen above that indeed the following rule is correct:

Trait C is selected if  $\frac{f_3}{p} > \frac{c}{d}$

For a given frequency, this indeed has the population structure comprised in a scalar— $f_3/p$ . This is also a proper function with the set of population structures as domain, because it only depends on  $f_0, f_1, f_2$  and  $f_3$ .

The reason why these two rules together constitute a counterexample against the existence of a rule that is (1) general, (2) correct and (3) uses a scalar representation of population structure is that these two use different scalar functions. In fact, it is sufficient to combine the set of linear public goods game with just a single game for which the rule is different; for instance Counterexample II, which is just one single game, would do. The essential point here is that there are numerous subsets of games for which there is a correct rule that has a scalar representing population structure in it. Each of those rules must imply that the set of games is reduced to a set of dimension 2. That can be done in different ways, each with its own function  $r(s)$ . But it is the fact that these functions  $r(s)$  are different that implies that a general rule with the same properties is not possible. Finding a subset of games for which the correct rule looks like  $r(s) > d(\pi)$  always implies a choice of a lower dimensional subset of the set of population structures  $S$  or—as in both examples here—of the set of games  $\Pi$ . In one subset the single parameter is  $r(s) = f_3/p$ , in the other it is  $r(s) = \mathbb{P}(T|T) - \mathbb{P}(T|N)$ . It is indeed true that a correct rule can consist of a function that compresses the population structure in a scalar, but we would have to find the right subset of



**Fig. 1.** A rule or prediction is general if it allows for any combination of choices from the set  $S$  of population structures and  $\Pi$  of games. With restrictions on the choice of population structures or games, one can arrive at rules in which the population structure is represented by a scalar. Two examples are games with generalized gains from switching, for which this scalar is  $r = \mathbb{P}(T|T) - \mathbb{P}(T|N)$ , and Marshall's set of stag-hunt games, for which it is  $f_3/p$ , but there are many more such subsets with ever differing ways to summarize the population structure by a scalar. There is however no general rule, as these scalar representations differ across subsets. Note that there is only one game in the intersection of Marshall's set of stag-hunt games and my set of games with generalized equal gains from switching; the game with all payoffs 0, in which strategies are indistinguishable and it is impossible for selection to act.

games that makes it correct, because how that scalar depends on the population structure differs from set to set (see Fig. 1).

That of course does not mean that I think inclusive fitness is useless. As a matter of fact, it could very well make sense to make restrictions. Maybe most real life situations only come from a small subset of population structures or a small subset of games. Maybe not. But whether or not realism would justify restrictions, it is clear that without them, there is no rule that satisfies payoff monotonicity, and in which population structure is reduced to a scalar for  $n > 2$ .

Then there is of course the possibility to stretch the concept of inclusive fitness, so that any condition of the form  $g(s, \pi) > 0$  is called a condition based on inclusive fitness. In the case of my group selection model that would be overstretching though, as the condition only entails the comparison of the average individual fitnesses of the different types.

In his comment, Marshall points to Queller (1992), saying that "Queller has shown that inclusive fitness models and group selection formulations both break down for the same reason when faced with non-additive fitness effects". What exactly that could mean is up for interpretation, but within the set of models of my paper, it is clear that a group selection model can be an accurate description of actual dynamics, while inclusive fitness can nonetheless give an inaccurate prediction. More precisely, with group size  $n > 2$ , we can conclude that a rule that reduces population structure to a scalar can only be correct for a subset of group selection models. That implies that such a rule produces inaccurate predictions for combinations of a population structure and a game that find themselves outside this subset. These combinations of population structures and games however still represent a model, which, if it matches observed dynamics, is not broken down.

## 6. Conclusion

There is a quote from Marshall's comment that provides a rather spot-on indication of what divides inclusive fitness optimists from inclusive fitness pessimists. It follows below.

Of course, it is now generally understood that the correct definition of relatedness is that which makes inclusive fitness theory work.

I would first like to point out—without saying whether I think that this claim is correct or not—that there is nothing "of course" about it. There is no question that relatedness matters, nor that it should be well defined, but making inclusive fitness work by choosing an appropriate definition of relatedness is not a priori

right or even possible. It could be the right thing to do, but it could also be that inclusive fitness is a Procrustean bed.<sup>3</sup> In order to indicate the subtleties involved, I would like to point at what I think is a particularly insightful analogy.

If the earth would be flat, then a 2-dimensional map would be a perfectly accurate model of the world. Also, if we would want to determine the shortest path between two places on earth, then it would naturally be a path that matches a straight line between them on the map.

The earth, however, is not flat, and the shortest path between two places across the surface of the earth is not necessarily a straight line on our standard map. If I fly from Amsterdam to Boston, then the path that the plane takes looks like an arc on the map. However, our standard map is not the only possible way to project our 3-dimensional earth on a 2-dimensional map. In fact, we can simply change the choice of the equator such that on the resulting map, the shortest distance between two particular cities is in fact a straight line.

If the question arises whether or not it is correct to claim that the shortest distance between two places on earth—across the surface—is a straight line on the map, then one possible answer is that indeed it is, if only we choose the appropriate map. That is true. It is also true that in order to determine what the appropriate map is, there is no way around processing the insight that the earth really is round and not flat. As a matter of fact, the most convenient way to create such a map is first to find the shortest distance on the globe and then to choose the equator such that the shortest path is part of it. It is a rather special way of finding a problem and a matching solution, but now the shortest path is indeed a straight line on the appropriate map.

Does this mean that 2-dimensional maps are useless for navigation? Of course not. A map can work perfectly well if we for instance assume that the two points, unlike Amsterdam and Boston, are just not too far apart. In that case the path with the shortest distance is very close to, or even indistinguishable from, a straight line on the (standard) map.

Making inclusive fitness work by choice of the appropriate definition of relatedness I think could be quite similar to finding the map to go with the shortest path. In this reply I have argued that within the simple group selection setting that I chose, there is a set of games of dimension 2 for which inclusive fitness works. We have also seen that there are actually many different sets of games of dimension 2 for which we can make a correct prediction based on

<sup>3</sup> Procrustes, a.k.a. "the stretcher" is a character in Greek mythology. He stretched people if they were too short to fit his bed, or cut off their legs if they were too long.

only a scalar that reflects population structure. What this scalar should be, however, differs from one set to the other, which makes relatedness the equivalent of the 2-dimensional map. This implies that if we take the set of games for which normal inclusive fitness works, and start increasing the dimension of the set of games we allow for, we will have to stretch the concept of inclusive fitness ever more in order to make it work. With every step towards generality, one has to adjust the inclusive fitness formula, in which we will need ever more numbers to represent the relevant features of the population structure, until at the end, when we arrive at a totally general prediction, the inclusive fitness formula is no different from a direct comparison of average individual fitnesses.

The point I try to make with the analogy is that even though inclusive fitness does not provide a generally correct prediction, I am not at all implying that I think inclusive fitness is useless. Also here, one can make additional assumptions under which even the simple, unadjusted inclusive fitness formula works. In the context of my group selection models such a condition is that the game has to satisfy generalized equal gains from switching. In other contexts other assumptions are possible. Inclusive fitness therefore may be quite useful, just as useful as a map is for finding the shortest path between two places that are not too far apart. But the point I wanted to make is that there are also situations where this is not the case.

**Acknowledgements**

I would like to thank James Marshall, Julian Garcia and the editor for useful comments, and an anonymous reviewer for especially precise and accurate suggestions for improvements.

**Appendix A. Translation Queller–van Veelen**

In this appendix, I would like to translate Eq. (3) in Queller (1985, p. 367) to Queller’s condition, as we derived it without the Price equation in Section 3, and which is stated in terms of my Counterexample I (van Veelen, 2009, pp. 594–595).

One thing we should realise is that in my model everything is deterministic, which implies that all covariances of all variables in the model are 0. That seems to be a dead end, but we can give the covariances in Queller an interpretation by thinking of a hypothetical chance experiment. Given a population that is characterized by the frequencies  $f_0, f_1$  and  $f_2$  we can imagine that one individual is drawn from the population, with all individuals having equal probability. The variances and covariances will then concern the phenotypic value of this randomly chosen individual and of its partner. In van Veelen (2009) the assortment parameter, or relatedness,  $r$  is defined with a similar reference to a hypothetical chance experiment.

In my model there is no difference between genotype and phenotype ( $G=P$ ), which implies that also the genotype can only take values 0 (if not altruistic) and 1 (if altruistic). This will simplify calculations significantly. Below I will give the probabilities that belong to different combinations of the randomly chosen individual ( $P$ ) and its partner ( $P'$ ). Here it is worthwhile to realise that the frequency  $p$ , the assortment parameter/relatedness  $r$ , and the different conditional probabilities are all functions of  $f_0, f_1$  and  $f_2$ . They relate in the following way: frequency  $p$  is total number of individuals with the gene divided by the total number of individuals. With a total of  $M$  groups, that would be  $(2f_2 + f_1)M/2M = f_2 + \frac{1}{2}f_1$ . This number is independent of  $M$ , so it is natural to also define  $p = f_2 + \frac{1}{2}f_1$  for infinite populations.

The variable  $r$  is defined in van Veelen (2009) as the difference between two probabilities in a general version of the hypothetical

chance experiment described above;  $r = \mathbb{P}(T|T) - \mathbb{P}(T|N)$ , where  $\mathbb{P}(T|T)$  here would translate to  $\mathbb{P}(P' = 1|P = 1)$ , and  $\mathbb{P}(T|N)$  to  $\mathbb{P}(P' = 1|P = 0)$ . The first one is the probability that your partner carries the trait if you do— $\mathbb{P}(T|T)$ —and here it is the number of individuals that carry the trait and that are matched with other carriers (both individuals in groups with 2 carriers, hence  $2f_2M$ ) divided by the total number of carriers of the trait, which is  $(2f_2 + f_1)M$ . This results in  $\mathbb{P}(T|T) = f_2/p$ . A similar computation gives us  $\mathbb{P}(T|N) = f_1/2(1-p)$ , which implies that  $r = (f_2/p) - f_1/2(1-p)$ . van Veelen (2009) also shows that with this definition of  $r$  we get  $\mathbb{P}(T|T) = r + (1-r)p$ , from which more or less directly follows that  $\mathbb{P}(N|N) = r + (1-r)(1-p)$ ,  $\mathbb{P}(N|T) = (1-r)(1-p)$  and  $\mathbb{P}(T|N) = (1-r)p$  (see also Grafen, 1985). So now we can write, for this hypothetical chance experiment:

$P$	$P'$	probability of this combination
1	1	$p \cdot \mathbb{P}(T T) = p(r + (1-r)p)$
1	0	$p \cdot \mathbb{P}(N T) = p(1-r)(1-p)$
0	1	$(1-p) \cdot \mathbb{P}(T N) = (1-p)(1-r)p$
0	0	$(1-p) \cdot \mathbb{P}(N N) = (1-p)(r + (1-r)(1-p))$

Note that, as we would expect,  $\mathbb{P}(10) = \mathbb{P}(01)$ . The expectations  $\mathbb{E}[P]$  and  $\mathbb{E}[P']$  are also what we would expect;  $\mathbb{E}[P] = p\mathbb{P}(T|T) + p\mathbb{P}(N|T) = p$  and  $\mathbb{E}[P'] = p\mathbb{P}(T|T) + (1-p) \cdot \mathbb{P}(T|N) = p$ .

With these probabilities we can compute the covariances in Queller’s (1985) rule. The first one is easy, again because  $P$  only takes values 0 and 1, which implies that  $P^2 = P$

$$\begin{aligned} Cov(P,P) &= Var(P) = \mathbb{E}[P^2] - \mathbb{E}^2[P] \\ &= \mathbb{E}[P] - \mathbb{E}^2[P] = p - p^2 = p(1-p) \end{aligned}$$

The second takes only a little more algebra:

$$\begin{aligned} Cov(P,P') &= \mathbb{E}[PP'] - \mathbb{E}[P]\mathbb{E}[P'] \\ &= p(r + (1-r)p) - p^2 = rp(1-p) \end{aligned}$$

The third is again simplified by  $P^2 = P$ :

$$\begin{aligned} Cov(P,PP') &= \mathbb{E}[P^2P'] - \mathbb{E}[P]\mathbb{E}[PP'] = \mathbb{E}[PP'] - \mathbb{E}[P]\mathbb{E}[PP'] \\ &= (1-p)p(r + (1-r)p) \end{aligned}$$

With this interpretation of the covariances, we can look at Queller’s equation (3), page 367, and see how it would translate. If we do, then we see that with the current interpretation of covariances as belonging to hypothetical chance experiments, and with genotype and phenotype always being equal, it becomes

Queller’s condition

$$\begin{aligned} -c + b \frac{Cov(P,P')}{Cov(P,P)} + d \frac{Cov(P,PP')}{Cov(P,P)} &> 0 \\ -c + b \frac{rp(1-p)}{p(1-p)} + d \frac{(1-p)p(r + (1-r)p)}{p(1-p)} &> 0 \\ -c + br + d(r + (1-r)p) &> 0 \end{aligned}$$

This matches the condition we derived in Section 3 without using the Price equation, assuming payoff monotonicity.

**Appendix B. Translation Marshall–van Veelen**

Given a population that is characterized by frequencies  $f_0, f_1, f_2$  and  $f_3$  we can, again, think of a hypothetical chance experiment. Imagine that one individual is drawn from the population, with all individuals having equal probability. Then we look at the phenotypic value of this individual and of its two partners. In my model there is still no difference between genotype and

phenotype ( $G=P$ ), which again implies that the genotype can only take values 0 (if not altruistic) and 1 (if altruistic). For our hypothetical chance experiment we can now write

$P$	$P'P''$	probability of this combination
1	11	$p \cdot \frac{3f_3}{3f_3+2f_2+1f_1} = f_3$
1	10 or 01	$p \cdot \frac{2f_2}{3f_3+2f_2+1f_1} = \frac{2}{3}f_2$
1	00	$p \cdot \frac{f_1}{3f_3+2f_2+1f_1} = \frac{1}{3}f_1$
0	11	$(1-p) \cdot \frac{f_2}{3f_0+2f_1+1f_2} = \frac{1}{3}f_2$
0	10 or 01	$(1-p) \cdot \frac{2f_1}{3f_0+2f_1+1f_2} = \frac{2}{3}f_1$
0	00	$(1-p) \cdot \frac{3f_0}{3f_0+2f_1+1f_2} = f_0$

Now we can compute the variance and covariance that feature in Marshall's equation. Again we use that  $P^2=P$

$$\begin{aligned} \text{Cov}(P, PP'P'') &= \mathbb{E}[P^2 P'P''] - \mathbb{E}[P] \mathbb{E}[PP'P''] = \mathbb{E}[PP'P''] - \mathbb{E}[P] \mathbb{E}[PP'P''] \\ &= (1-p) \mathbb{E}[PP'P''] = (1-p)f_3 \end{aligned}$$

The variance is standard:

$$\text{Var}(P) = p(1-p)$$

That implies that if we reinterpret Marshall's condition with the hypothetical chance experiment described above, we arrive at the same condition that is derived above:

Marshall's condition

$$\frac{\text{Cov}(P, PP'P'')}{\text{Var}(P)} > \frac{c}{d}$$

$$\frac{(1-p)f_3}{p(1-p)} > \frac{c}{d}$$

$$\frac{f_3}{p} > \frac{c}{d}$$

This matches the condition we derived in Section 4, without using the Price equation, assuming payoff monotonicity.

It is worth noting that payoff monotonicity and Marshall's rule are equivalent for this game, as payoff monotonicity and Queller's

rule were for the game from Section 3. Although payoff monotonicity is a very mild assumption, it is an assumption nonetheless, and this shows that the derivation with the Price equation obscures the fact that there is an assumption behind the rule.

## References

- Grafen, A., 1985. A geometric view of relatedness. *Oxford Surveys in Evolutionary Biology* 2, 28–90.
- Grafen, A., 2007. Detecting kin selection at work using inclusive fitness. *Proc. R. Soc. B* 274, 713–719.
- Hofbauer, J., Sigmund, K., 1998. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK.
- Hamilton, W.D., 1964a. The genetical theory of social behaviour (I and II). *J. Theor. Biol.* 7 1–16.
- Hamilton, W.D., 1964b. The genetical theory of social behaviour II. *J. Theor. Biol.* 7, 17–32.
- Killingback, T., Bieri, J., Flatt, T., 2006. Evolution in group-structured populations can resolve the tragedy of the commons. *Proc. R. Soc. B* 273, 1477–1481.
- Lehmann, L., Keller, L., West, S.A., Roze, D., 2007. Group selection and kin selection: two concepts but one process. *Proc. Natl. Acad. Sci.* 104, 6736–6739.
- Marshall, J.A.R. Queller's rule ok: Comment on van Veelen 'when inclusive fitness is right and when it can be wrong'. *J. Theor. Biol.*, this issue, doi:10.1016/j.jtbi.2010.07.015.
- Nowak, M.A., Tarnita, C.E., Wilson, E.O., 2010. The evolution of eusociality. *Nature* 466, 1057–1062.
- Queller, D.C., 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* 318, 366–367.
- Queller, D.C., 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139 (3).
- Traulsen, A., Nowak, M., 2006. Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci.* 103, 10952–10955.
- van Veelen, M., 2005. On the use of the Price equation. *J. Theor. Biol.* 237, 412–426.
- van Veelen, M., 2007. Hamilton's missing link. *J. Theor. Biol.* 246, 551–554.
- 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.
- Vega-Redondo, F., 2003. *Economics and the Theory of Games*. Cambridge University Press, Cambridge, UK.
- Wade, M.J., Wilson, D.S., Goodnight, C., Taylor, D., Bar-Yam, Y., de Aguiar, M.A.M., Stacey, B., Werfel, J., Hoelzer, G.A., Brodie III, E.D., Fields, P., Breden, F., Linksvayer, T.A., Fletcher, J.A., Richerson, P.J., Bever, J.D., Van Dyken, J.D., Zee, P., 2009. Multilevel and kin selection in a connected world. *Nature* 463, E8–E9.
- Wild, G., Gardner, A., West, S.A., 2009. Adaptation and the evolution of parasite virulence in a connected world. *Nature* 459, 983–986.

Matthijs van Veelen \*

CREED, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB  
Amsterdam, the Netherlands  
E-mail address: C.M.vanVeelen@uva.nl

Available online 13 November 2010