SUPPLEMENTARY INFORMATION

Contents

Part A – Natural selection versus kin selection

Kin selection theory based on the concept of inclusive fitness is often presented as a general approach that can deal with many aspects of evolutionary dynamics. Here we show that this is not the case. Instead, inclusive fitness considerations rest on fragile assumptions, which do not hold in general. The quantitative analysis of kin selection relies completely on inclusive fitness theory. No other theory has been proposed to discuss kin selection. We show the limitations of inclusive fitness theory. We do not discuss implications of kin selection that might exist independent of inclusive fitness theory.

We set up a general calculation for analyzing mutation and selection of two strategies, A and B, and then derive the fundamental condition for one strategy to be favored over the other. This condition holds for any mutation rate and any intensity of selection. Subsequently, we limit our investigation to weak selection, because this is the only ground that can be covered by inclusive fitness theory. For weak selection, we show that the natural selection interpretation is appropriate for all cases, whereas the kin selection interpretation, although possible in several cases, cannot be generalized to cover all situations without stretching the concept of "relatedness" to the point where it becomes meaningless.

Therefore we have a general theory, based on natural selection and direct fitness, and a specific theory based on kin selection and inclusive fitness. The general theory is simple and covers all cases, while the specific theory is complicated and works only for a small subset of cases. Whenever both theories work, inclusive fitness does not provide any additional insights. Criticisms of inclusive fitness theory have already been raised by population geneticists and mathematicians (Cavalli-Sforza and Feldman 1978 and Karlin and Matessi 1983). The present criticism is based on a game theoretic perspective in structured populations which has been developed recently.

The extra complication of inclusive fitness theories arises from the attempt to bring into the discussion increasingly abstract notions of 'relatedness' when it is not natural to do so. This situation is not particular to theory. Hunt (2007) citing Mehdiabadi et al (2003) points out that "increasingly complex scenarios are required to keep recent empirical data within the theoretical construct of haplodiploidy-based maximization of inclusive fitness."

The fact that inclusive fitness calculations are more complicated than direct fitness calculations has been accepted by theoreticians such as Rousset and Billiard (2000) and Taylor et al (2006). As complicated as inclusive fitness is to calculate, it is even more complicated to measure empirically. Only very few studies have attempted to do this (Queller and Strassman 1989, Gadagkar 2001) and their results have been mixed. Despite the difficulty of measuring inclusive fitness, it is often possible to measure genetic relatedness, which has acted as an endorsement for inclusive fitness theoreticians. However, measuring relatedness (instead of inclusive fitness) can lead to misleading results: after getting recognition from proposing that haplodiploidy is the reason for insect sociality, Hamilton's rule has lost steam when many studies have shown that there is in fact no apparent link between the two (Anderson 1984, Gadagkar 1991, Crozier and Pamilo 1996, Queller and Strassmann 1998, Linksvayer and Wade 2005, Hunt 2007, Boomsma 2009). In Section 6, we also give a simple example to show that relatedness measurements, in the absence of a model, can be very misleading.

Those who have attempted thorough assessments of inclusive fitness have come to the similar conclusion that "ecological, physiological, and demographic factors can be more important in promoting the evolution of eusociality than the genetic relatedness asymetries" (Gadagkar 2001). In other words, a thorough understanding of the many factors at play is much more important than the isolated measurement of relatedness. We aim to show that the understanding of such factors would lead to the design of solid models. When such models are proposed and analyzed using natural selection, then measurements of genetic relatedness could receive meaningful interpretations.

We fail to see the point in insisting to assign explanatory power to a theory which from a modeling perspective fails to cover the majority of cases (and where it does, it makes the same predictions as natural selection) and which moreover has limited support in the empirical world.

We recognize that inclusive fitness theory has led to important findings such as the elegant framework proposed by Rousset & Billiard (2000) and Roze & Rousset (2004), the results of Taylor (1989) regarding evolutionary stability in one-parameter models (with some improvements proposed by van Veelen 2005) and the results of Taylor et al (2007a) for homogeneous graphs . But on the other hand many recent contributions of inclusive fitness theory consist of either rederiving special cases of known results (Lehmann et al 2007ab, Taylor and Grafen 2010) or of making incorrect universality claims (Lehmann and Keller 2006, West et al 2007, Gardner 2009, West and Gardner 2010).

In light of what we show here, inclusive fitness theory is simply a method of calculation, but one that works only in a very limited domain. Endorsing it as a mechanism for the evolution of cooperation would lead to a constraining view of the world (also pointed out in Nowak et al 2010). Hunt (2007) says that Hamilton's rule, proposed as a general rule with broad explanatory power, "has blunted inquiry into mechanisms that foster and maintain sociality in the diverse lineages where sociality has evolved." Similarly, from a theoretical perspective, the narrow focus on relatedness has prevented kin selectionists from contributing to the discovery of mechanisms for the evolution of cooperation. Such mechanisms lead to an assortment between cooperators and defectors, but assortment itself is not a mechanism; it is the consequence of a mechanism. The crucial question is always how assortment is achieved (Nowak et al 2006).

1 Mutation-selection analysis

We consider stochastic evolutionary dynamics (with mutation and selection) in an asexual population of finite size, N. We do not specify yet the underlying stochastic process because our results are general and apply to a large class. Individuals adopt either strategy A or B. They obtain payoff by interacting with others according to the underlying process. This payoff determines the reproductive success of an individual. We call this the 'natural selection approach'.

Reproduction is subject to mutation. With probability u the offspring adopts a random strategy (which is either A or B). With probability $1-u$ the offspring adopts the parent's strategy. Thus, mutation is symmetric and occurs during reproduction.

As a consequence of the underlying dynamics, the process goes through many states. Each state, S, is a snapshot of the process and is described by the strategies of all individuals (A or B) as well as by their 'locations' (in space, phenotype space, on islands, on sets, etc). A description of a state must include all information that is necessary to obtain the payoffs of individuals in that state. For our discussion, we assume a finite state space, but the analysis can be extended to infinite state spaces. We study a Markov process on this state space.

One could ask many questions about such a system. Does selection lead to dominance, bistability or coexistence? What are the trajectories of the system? What is the stationary distribution? And so on. These are all questions concerning the dynamics. The stochastic element of evolution, which leads to a distribution of possible outcomes rather than a single optimum, is not a part of inclusive fitness theory, while it is essential to evolutionary genetic theory. Inclusive fitness theory can only attempt to address two types of questions, both of them insufficient to analyze the whole dynamics. First, it can determine whether cooperation is favored by looking at the gradient of selection. However, as it has already been pointed out, this measure only works if selection is not frequency dependent. In other words, it works only when fitness gradients are determined entirely by processes that are not affected by the current state of the population (Doebeli and Hauert 2006, Traulsen 2010). Moreover, Wolf and Wade (2001) have shown that the inclusive fitness approach of counting offspring viability as a component of maternal fitness can lead to a mistaken understanding of the direction of selection. Since the limitations of such a method are clear and have already been pointed out carefully, we will not deal with this type of question here. The second question that inclusive fitness can attempt to answer has to do with determining which strategy is more abundant on average in the stationary distribution.

A natural selection approach is from the beginning broader than the inclusive fitness approach because it can handle questions about dynamics (Traulsen 2010). But since in this paper we are aiming to compare the natural selection approach to the inclusive fitness approach, we will only address the question that can be answered by the latter: when is one strategy more abundant than another on average?

The system goes through many states, and some states are less visited than others. We follow the process over many generations; in some states A players do better, in others they do worse. For the purpose of this analysis, all that matters is how they fare on average. We say that on average A outperforms B if the average frequency of A is greater than $1/2$. Let x_S be the frequency of A in state S. Then A is favored over B on average if

$$\langle x \rangle = \sum_S x_S \pi_S > \frac{1}{2}. \tag{1}$$

Here $\langle \cdot \rangle$ denotes the average taken over the stationary distribution and π_S is the probability to find the system in state S (or, in other words, the fraction of time spent by the system in state S). In the limit of low mutation, this condition is equivalent to the comparison of fixation probabilities, $\rho_A > \rho_B$.

To tackle this problem, given the general process described above, we can write an intuitive description of how the frequency of A or B changes from one state to another. This type of argument has been used several times, starting with Price (1970, 1972), who used it for processes with non-overlapping generations. It has been more carefully revised by Rousset and Billiard (2000)

for simple deme structures. The same type of analysis has been employed for games in phenotype space (Antal et al 2009) and for games on sets (Tarnita et al 2009a). None of these accounts deal with general processes. But in what follows we give a general mutation-selection analysis, which does not assume a particular process or dynamics. Moreover, we do not specify how selection plays a role in the process.

To understand how the frequency of A changes between states, we must take into account the two forces that act: selection and mutation. In the stationary distribution, mutation and selection balance each other on average. Hence the total change in the frequency of A is zero, when averaged over the stationary distribution:

$$0 = \langle \Delta x^{tot} \rangle \tag{2}$$

From now on, whenever we write the stationary average of a quantity, we use the angular brackets $\langle \cdot \rangle$; however, when we refer to quantities in a state, we omit, for simplicity, the index S. The indication that we refer to the quantity in a state rather than to its average over the stationary distribution comes from the fact that we do not use the angular brackets for the former.

Let w_i denote the expected fitness of individual i. As mentioned, this quantity is for a given state, hence the lack of angular brackets. We can decompose w_i into two parts. One is the expected number of offspring, b_i, and the other is the expected number of survivors, $1-d_i$, where d_i represents the probability that i dies in a selection step. Thus, the expected fitness of individual i is

$$w_i = 1 - d_i + b_i \tag{3}$$

Since the population size is fixed, we have $\sum_i w_i = N$, which implies $\sum_i b_i = \sum_i d_i$.

In a given state, the total expected change in the frequency of A can be expressed in terms of birth and death rates as follows. There are two ways to produce more A individuals: the existing ones give birth and their offspring do not mutate to B or the existing B individuals give birth and their offspring mutate to A. There is however only one way to lose A, and this is if some existing A individuals die. Thus, in a given state, the total change in frequency due to selection and mutation is

$$\Delta x^{tot} = \frac{1}{N}\left(\left(1-\frac{u}{2}\right)\sum_i s_i b_i + \frac{u}{2}\sum_i (1-s_i)b_i - \sum_i s_i d_i\right). \tag{4}$$

Here s_i indicates the strategy of individual i: $s_i = 1$ if i has strategy A and it is 0 if i has strategy B.

On the other hand, the change only due to selection is simply the expected number of offspring of A individuals minus the number of A in this state:

$$\Delta x^{sel} = \frac{1}{N}\sum_i s_i(w_i - 1) = \frac{1}{N}\sum_i s_i(b_i - d_i). \tag{5}$$

Using (5) into (4) together with the fact that $Nx_S = \sum_i s_i$ we can rewrite the total change in terms of the change due to selection[1]

$$\Delta x^{tot} = \Delta x^{sel} + \frac{u}{2N}\sum b_i - \frac{u}{N}x - \frac{u}{N}\sum_i s_i\Big(b_i - \frac{1}{N}\Big). \tag{6}$$

[1]This way of writing it is in no way unique but any other rewriting will yield the same results.

This type of accounting analysis is a generalization of Price's (1970, 1972) method and agrees with it for a process with non-overlapping generations. However, we keep our result in the form of this accounting identity and do not use notations like covariance which have been shown to be confusing if used to make predictions in the absence of a precise model (as explained by van Veelen 2005).

Next we look at average quantities (similar to Billiard and Rousset 2000, Antal et al 2009a, Tarnita et al 2009a). Since the total change averaged over the stationary distribution is zero, we have

$$0 = \langle \Delta x^{tot} \rangle = \langle \Delta x^{sel} \rangle + \frac{u}{2N} \Big\langle \sum b_i \Big\rangle - \frac{u}{N} \langle x \rangle - \frac{u}{N} \Big\langle \sum_i s_i \Big(b_i - \frac{1}{N} \Big) \Big\rangle. \tag{7}$$

Thus, we can rewrite the average frequency in terms of the average change due to selection as

$$\langle x \rangle = \frac{1}{2} \Big\langle \sum b_i \Big\rangle + \frac{N}{u} \langle \Delta x^{sel} \rangle - \Big\langle \sum_i s_i \Big(b_i - \frac{1}{N} \Big) \Big\rangle \tag{8}$$

We want to compare the average frequency of A to $1/2$. For simplicity, we make the following assumption.

> **Assumption (1).** The total birth rate (or, equivalently, the total death rate) is the same in every state.

In other words, we assume $\sum_i b_i = \alpha$ in all states of the system, where α is some constant. This assumption is not as restrictive as it may seem. It holds for most processes that have been analyzed so far. It holds for Wright-Fisher type processes. There, all individuals from one generation die, and the new generation is formed by their offspring. Thus, the death rate of each individual is $d_i = 1$ and so $\sum_i d_i = N = \sum_i b_i$ (here $\alpha = N$).

Assumption (1) also holds for Moran type processes (Moran 1962) with either Death-Birth (DB) or Birth-Death (BD) updating (Ohtsuki et al 2006, Ohtsuki & Nowak 2006). DB updating means that an individual dies at random and others compete for the empty site proportional to their payoff. We have $d_i = 1/N$ for all i yielding that $\sum_i d_i = 1 = \sum_i b_i$. BD updating means that an individual is chosen for reproduction proportional to payoff and the offspring replaces a randomly chosen individual. In this case we have $b_i = f_i/F$ where F is the total payoff in the population and therefore $\sum_i b_i = \sum_i f_i/F = 1$. For Moran type processes $\alpha = 1$.

The derivation works for any constant α but for simplicity of exposition we set $\alpha = 1$. Then if for every state $\sum_i b_i = 1$, we have $\langle \sum_i b_i \rangle = 1$. Hence (8) becomes

$$\langle x \rangle = \frac{1}{2} + \frac{N}{u} \langle \Delta x^{sel} \rangle - \Big\langle \sum_i s_i \Big(b_i - \frac{1}{N} \Big) \Big\rangle \tag{9}$$

Strategy A is favored over B if $\langle x \rangle > 1/2$. Therefore, we obtain the main result

Theorem 1. *For any process satisfying assumption (1) and for any intensity of selection, A is favored over B in the mutation-selection equilibrium if and only if*

$$\Big\langle \sum_i s_i (b_i - d_i) \Big\rangle > u \Big\langle \sum_i s_i \Big(b_i - \frac{1}{N} \Big) \Big\rangle \tag{10}$$

If the birth rate is constant, since $\sum_i b_i = 1$, we must have $b_i = 1/N$ and then condition (10) reduces to $\langle \Delta x^{sel} \rangle > 0$. It is easily shown that the same is true if the death rate is constant. This result is valid for any mutation rate. If however we consider the limit of low mutation in (10), for any process, we recover the same condition $\langle \Delta x^{sel} \rangle > 0$. We can then formulate the following

Corollary 1. *If we consider either constant birth or constant death rates or if we consider the limit of low mutation, A is favored over B if*

$$\langle \Delta x^{sel} \rangle > 0$$

This result is intuitive, because in the limit of low mutation, only selection determines whether a strategy is favored or not. However, in general, the condition for a strategy to be favored has to take into account both selection and mutation. If selection favors a strategy, then it might reproduce more often which makes it subject to mutation. Thus, in this case, it is not enough to ask for selection to favor a strategy. One must require that selection favors the strategy enough to offset the counter effect of mutation. This argument explains (10); the right hand side of (10) involves only the birth rate multiplied by the mutation probability.[2]

So far we have derived a general condition for one strategy to be favored over another in a mutation selection process. We did not need to specify the process in detail. Moreover, our result holds for any intensity of selection. Next we focus on the limit of weak selection because this is the only ground covered by inclusive fitness theory.

2 The limit of weak selection

Inclusive fitness theory works only for the limit of weak selection (Michod and Hamilton 1980, Grafen 1984). In this limit the selective difference between the two strategies converges to zero. Therefore both strategies have frequency of about 1/2 with an epsilon difference determining the winner. The weak selection limit is a useful, simplified scenario for gaining some insights into evolutionary dynamics, but it is obviously not the general case. Therefore, if a theory, such as inclusive fitness theory, can only be formulated for weak selection, it cannot possibly represent a general principle of evolutionary biology.

2.1 Two types of weak selection

The limit of weak selection can be achieved in different ways. Here we discuss two possibilities. For the approach proposed by Nowak et al (2004), the intensity of selection scales the contribution of the game relative to the baseline payoff. Thus, the effective payoff of an individual is $1 + w\text{Payoff}$. The limit of weak selection is obtained for $w \to 0$.

In inclusive fitness theory, weak selection is obtained by assuming that mutation renders strategies which are very close to the wild type in phenotype space (Taylor 1989, Rousset and Billiard 2000). These two papers deal with a continuous strategy space, but it seems to be very common in inclusive fitness theory to assume that the strategy space is the space of mixed strategies, which is a special case (Grafen 1979, Wild and Traulsen 2007, Traulsen 2010). In this case, weak selection

[2]If instead of assuming $\sum_i b_i = 1$ we would assume $\sum_i b_i = \alpha$ then (10) would become $\left\langle \sum_i s_i(b_i - d_i) \right\rangle > u \left\langle \sum_i s_i \left(b_i - \frac{\alpha}{N} \right) \right\rangle$. This is easily verified and the entire analysis carries out similarly.

corresponds to small deviations in the probability to play a certain strategy. More explicitly, in this special case, weak selection is obtained as follows. If a game is given by two pure strategies, X and Y, then one considers the set of all mixed strategies given by the probability p to play X (the probability to play Y is $1-p$). Then one studies selection between the two strategies, p and $p+\delta$. The limit of weak selection is given by $\delta \to 0$. The payoff matrix for the game between these two strategies, p and $p+\delta$, in the limit of weak selection, has the property of "equal gains from switching", which means that the sum of the entries on the first diagonal equals the sum of the entries on the second diagonal (Nowak and Sigmund 1990). Thus, this limit of weak selection leads to the necessary constraint that the games have equal gains from switching. For more details on these two types of weak selection we refer to Rousset and Billiard (2000), Wild and Traulsen (2007) and Traulsen (2010).

In what follows, we use δ to denote the intensity of selection, but our theory holds for both approaches to weak selection.

2.2 Weak selection of strategies

As mentioned, (10) holds for any intensity of selection. It holds for both selection approaches described above, as well as for any other suggestions of how selection should play a role in the model. In this section we consider that the intensity of selection is specified by a parameter δ but we do not yet specify how the parameter δ plays a role. Hence the derivation below is for weak selection in the most general form and it holds provided that there is no discontinuity between neutrality and selection; thus we assume that all quantities that depend on δ are differentiable at $\delta = 0$.

In a selection based approach, the birth and death rates of individuals depend on the parameter δ. For the limit of weak selection, $\delta \to 0$, we can take the Taylor expansion at $\delta = 0$. From (10) we obtain

$$\Big\langle \sum_i s_i(b_i - d_i) \Big\rangle_0 + \delta \frac{\partial}{\partial \delta} \Big\langle \sum_i s_i(b_i - d_i) \Big\rangle \Big|_{\delta=0} > u\Big(\Big\langle \sum_i s_i\Big(b_i - \frac{1}{N}\Big) \Big\rangle_0 + \delta \frac{\partial}{\partial \delta} \Big\langle \sum_i s_i\Big(b_i - \frac{1}{N}\Big) \Big\rangle \Big|_{\delta=0} \Big) \tag{11}$$

Here $\langle \cdot \rangle_0 = \sum_S \cdot \pi_S \mid_{\delta=0}$ denotes the average over the stationary distribution taken at neutrality and $\mid_{\delta=0}$ means the quantity is evaluated at $\delta = 0$. This can be expanded further as

$$\begin{aligned} &\Big\langle \sum_i s_i(b_i - d_i) \Big\rangle_0 + \delta \Big(\sum_S \sum_i s_i(b_i - d_i) \mid_{\delta=0} \frac{\partial}{\partial \delta} \pi_S \Big|_{\delta=0} + \sum_S \frac{\partial}{\partial \delta} \sum_i s_i(b_i - d_i) \Big|_{\delta=0} \pi_S \mid_{\delta=0} \Big) \\ &> u \Bigg(\Big\langle \sum_i s_i\Big(b_i - \frac{1}{N}\Big) \Big\rangle_0 + \delta \Big(\sum_S \sum_i s_i\Big(b_i - \frac{1}{N}\Big) \mid_{\delta=0} \frac{\partial}{\partial \delta} \pi_S \Big|_{\delta=0} + \sum_S \sum_i \frac{\partial}{\partial \delta} s_i\Big(b_i - \frac{1}{N}\Big) \Big|_{\delta=0} \pi_S \mid_{\delta=} \end{aligned} \tag{12}$$

Computationally it is easy to deal with averages over the neutral stationary distribution, because they allow us to calculate population structure at neutrality. Thus, in the above expression the first and last terms can be calculated. The problem arises with the middle terms of the form

$$\sum_S s_i(b_i - d_i) \mid_{\delta=0} \frac{\partial}{\partial \delta} \pi_S \Big|_{\delta=0} \tag{13}$$

In such terms, the quantities are still evaluated at neutrality, but we have to calculate the first derivative of the actual stationary distribution for any intensity of selection, and then evaluate that at $\delta = 0$. It is usually very hard to find the stationary distribution for any δ (if that could be done, there would be no need to take the limit of weak selection). Hence terms like (13) would be hard to handle unless they are zero. One way to make such terms zero is via the following assumption.

Assumption (2). At neutrality ($\delta = 0$), all birth rates and all death rates are equal in every state.

By this we mean that $b_i = d_i = 1/N$ for al i, in every state[3].[4] If this is the case, then $(b_i - 1/N) \mid_{\delta=0} = (b_i - d_i) \mid_{\delta=0} = 0$. Moreover, we implicitly get that $\langle \Delta x_{sel} \rangle_0 = 0$ and $\langle \sum_i s_i (b_i - 1/N) \rangle_0 = 0$. Then, (12) simplifies giving the equivalent of our main result (10) for weak selection:

Theorem 2. *For any process satisfying assumptions (1) and (2), in the limit of weak selection, strategy A is favored over strategy B in the mutation-selection equilibrium if and only if*

$$\left\langle \sum_i s_i \left. \frac{\partial (b_i - d_i)}{\partial \delta} \right|_{\delta=0} \right\rangle_0 > u \left\langle \sum_i s_i \left. \frac{\partial b_i}{\partial \delta} \right|_{\delta=0} \right\rangle_0 \tag{14}$$

Corollary 2. *In particular, if we also consider either constant birth or constant death rates or if we consider the limit of low mutation,* (14) *becomes*

$$\left\langle \sum_i s_i \left. \frac{\partial (b_i - d_i)}{\partial \delta} \right|_{\delta=0} \right\rangle_0 > 0 \tag{15}$$

3 Comparing natural selection and kin selection

Our main result (Theorem 1) was derived for any intensity of selection. Subsequently we took the limit of weak selection to derive the more convenient condition (Theorem 2). But the standard approach of natural selection is, of course, not limited to weak selection. We recognize that many interesting and important phenomena of evolutionary dynamics can only be observed if we move away from the limit of weak selection (since for weak selection all available strategies are equally abundant on average).

The whole theory of inclusive fitness, however, is only applicable in the limit of weak selection (Michod and Hamilton 1980, Grafen 1984). The fundamental idea of inclusive fitness is that the consequence of an action can be evaluated as the sum of the following terms: the fitness effect that this action has on the actor plus the fitness effect that this action has on any recipient multiplied by the relatedness between actor and recipient. This is a somewhat artificial and tricky construct that has confused people. According to Grafen (1984) many erroneous definitions are given in

[3] Or in the more general case, $b_i = d_i = \alpha/N$.

[4] One immediate example for which $b_i \neq d_i$ is the star. The star is a graph with a hub and $N-1$ leaves. Consider DB updating on the star – if a leaf dies then the hub will replace it; if the hub dies, then the leaves compete for the empty spot. At neutrality, each individual has payoff 1. Then $d_i = 1/N$ for all i; $b_i = 1/N(N-1)$ for the leaves and for the hub $b_{hub} = (N-1)/N$. So clearly $b_i \neq d_i$ for all i. Thus, for the star, the analysis even for weak selection is more complicated. Results are possible using our natural selection based approach but only for low mutation (Tarnita et al 2009b).

textbooks and then used for subsequent theoretical or empirical studies. The important point is that inclusive fitness does not count the offspring of an individual that come from others' actions on him; so it does not include the whole classical fitness of an individual (see Figure 1). The suggestion of inclusive fitness theory is that this (somewhat artificial) construct can be used to evaluate evolutionary dynamics. In the following, we show that the concept of inclusive fitness only makes sense if several very restrictive assumptions hold (in addition to the already very restrictive assumption of weak selection).

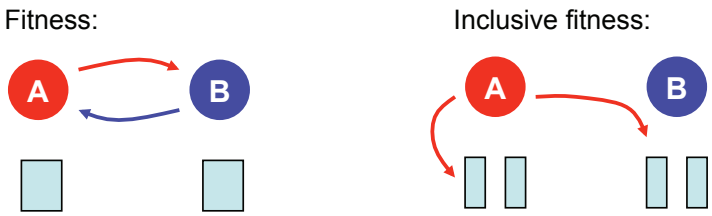

Figure 1: **Inclusive fitness is simply a different accounting method that works in some cases, but when it works it never has an advantage over the standard fitness concept of natural selection.** For calculating the fitness of an individual we consider all interactions and then calculate how the payoff is translated into reproductive success. Inclusive fitness is the sum of how the action of an individual affects his own fitness plus how this action affects the fitness of another individual multiplied by the relatedness between the two. Inclusive fitness does not take into account the fitness contribution that arises from the action of others on the focal individual.

For instance, for inclusive fitness to work, one has to assume that the effects of one's behavior on others are linear, additive and independent. In other words, for calculating inclusive fitness, it must be sufficient to look at pairwise interactions independently, and such interactions can then be added up. If stronger selection or synergistic effects are at work, an expression of the form (16) (shown below) cannot be written anymore. We will show these and other failures of the inclusive fitness concept in Section 7.

For non-vanishing intensity of selection, it matters how selection is incorporated into the model, whether it affects the payoff entries, whether it reflects distance in phenotype space and so on. However, we will show below that in the limit of weak selection and under certain additional assumptions, (10) will yield the same result for at least the above two types of selection. This is where the debate arises – under certain assumptions and for weak selection, both the inclusive fitness and the natural selection approaches are identical. However, as one moves away from weak selection or if these simplifying assumptions are not fulfilled, the inclusive fitness approach cannot be generalized further without making it so contrived that it loses its meaning. In these circumstances, the natural selection approach is the *natural* approach to be employed. And hence one may argue that if you have a theory that works for all cases (natural selection) and a theory that works for only some cases (kin selection) and where it works, it agrees with the general theory, why not simply use the general theory everywhere?

Hamilton's (1964) paper provides the framework for the inclusive fitness approach. For a recent and thorough discussion of Hamilton's central results, see van Veelen (2007). The formulation of inclusive fitness that we use below is the one that is currently used in inclusive fitness theory (Taylor and Frank 1996, Taylor et al 2007b). A focal A-individual (the actor) is chosen, which is representative of the average. Its strategy is $s_\bullet$ and its fitness is $w_\bullet = 1 - d_\bullet + b_\bullet$. Then the effects of its A-behavior on all individuals in the population (the recipients) are added, each effect weighted by the "relatedness" R of the actor to the recipient. The inclusive fitness effect of the focal individual is then written in the limit of weak selection and low mutation as[5]

$$W_{IF} = \sum_j \frac{\partial}{\partial \delta} \frac{\partial w_j}{\partial s_\bullet}\bigg|_{\delta=0} R_j. \tag{16}$$

Here R_j is the relatedness of the focal individual to recipient j. For the asexual model that we are interested in, it is defined as

$$R_j = \frac{Q_j - \overline{Q}}{1 - \overline{Q}} \tag{17}$$

where $Q_j = \Pr(s_\bullet = s_j) = 2\langle s_\bullet s_j \rangle_0 / N$ is the probability that the focal individual and individual j are identical by descent (IBD), at neutrality, and $\overline{Q}$ is the average identity by descent. Note that this definition is not given in a state; it is given on average. If we were to give the definition in a state, the average $\langle s_\bullet s_j \rangle_0$ would have to be replaced by $s_\bullet s_j$ calculated in that particular state. This latter quantity is determined by the labels of the two players in the given state and has nothing to do with identity by descent or with relatedness. Hence, in order to have an inclusive fitness definition that has relatedness in it, one needs to define inclusive fitness as an average. This is then inherently different from usual fitness, which is defined in a state.

Note moreover that what is called "relatedness" by theoreticians is not a measure of genetic identity (that would be Q_j) but a measure of relative genetic identity. Due to this normalization, the relatedness of any individual to oneself is one and the relatedness to the population is zero. A consequence of the latter is that the actor also has negative "relatedness" to some fraction of the population. Inclusive fitness theory focuses on low mutation, hence the only interesting effect is that due to selection.

Inclusive fitness theory then says that strategy A is favored over strategy B if

$$W_{IF} > 0 \tag{18}$$

Our main result for weak selection and low mutation (15) does not make any assumptions about the role 'relatedness' plays in the model. It is a general condition and below we show under which assumptions (15) can be reduced to $W_{IF} > 0$.

First, we point out that in the kin selection literature, it has already been shown that calculating inclusive fitness is more cumbersome than calculating direct fitness. Taylor et al (2006) write "direct fitness can be mathematically easier to work with and has recently emerged as the preferred approach of theoreticians." By direct fitness, kin selection theoreticians mean looking at the effects

[5]Without trying to be pedantic, we would like to point out that this notation is very unfortunate. The symbol ∂ denotes differentiation but phenotypes might be discrete and not continuous variables, so differentiation with respect to them (as in $\partial w_j / \partial s_\bullet$) does not make sense. We only reproduce it here for historical purposes, as this is the way it has been used in the inclusive fitness literature for decades.

of everyone in the population on a given recipient[6], and weighing those effects by the relatedness between each actor and the recipient

$$W_{dir} = \sum_j \frac{\partial}{\partial \delta} \frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} R_j \tag{19}$$

This is simply a reformulation of inclusive fitness (18) in terms of direct fitness. It is easily shown that $W_{IF} > 0$ is equivalent to $W_{dir} > 0$ (Rousset 2004, Taylor et al 2006).

We are advocates of direct fitness as well (albeit in a more general form than W_{dir}) and appreciate the fact that kin selection theoreticians start to employ direct fitness methods. However, our direct fitness method based on our main result (15) is more general than (19) simply because we do not constrain ourselves to a method that must weigh effects by relatedness – we let the model guide us as to how the direct effects play a role. Below we aim to show under what assumptions our weak selection condition (15) is equivalent to (19) and implictly to (18).

3.1 Additional assumptions needed for inclusive fitness theory

The following assumptions are necessary for inclusive fitness theory to be defined and to work.

Assumption (i). The game is additive.

This means that all interactions between individuals occur pairwise and the effects of all such pairwise interactions can be added up to determine an individual's overall payoff. In Section 7 we discuss what happens if this assumption fails. For now we focus on what happens when this assumption holds.

If the game is additive, we can express (15) as

$$\left\langle \sum_i s_i \sum_j \frac{\partial}{\partial \delta} \frac{\partial w_i}{\partial s_j}\bigg|_{\delta=0} s_j \right\rangle_0 = \left\langle \sum_i \sum_j \frac{\partial}{\partial \delta} \frac{\partial w_i}{\partial s_j}\bigg|_{\delta=0} s_i s_j \right\rangle_0 > 0 \tag{20}$$

Although summing over all individuals is the more accurate way to do it, one could also, given a symmetry of the individuals[7], choose a focal individual $\bullet$ representative of the average and rewrite the above condition as

$$\left\langle \sum_j \frac{\partial}{\partial \delta} \frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} s_\bullet s_j \right\rangle_0 > 0 \tag{21}$$

This is closer but still quite different from (19). Here the average is still taken over all elements of the sum, rather than simply over the strategies. Condition (21) is more general than (19), because it assumes (correctly) that the fitness of individuals depends on the interaction structure which can vary between states. In that case, one cannot separate the effect of the structure from that of "relatedness" as we will show in Section 7. One can only rewrite (21) as (19) if the following assumption also holds

[6] As opposed to an inclusive fitness approach where one would trace the effects of an actor on everyone else in the population.

[7] This assumption can also easily fail, but here we will not address this issue because it is too technical.

Assumption (ii). The population structure is 'special' (non-generic).

A 'special' population structure satisfies one of the following two criteria:

(iia) The population structure is static.

(iib) The population structure is dynamic, but in the restricted way that two individuals either interact or they do not interact (which means interaction is all or nothing) and the updating is global (which means everyone competes globally with everyone else for reproduction).

Examples of population structures that fulfill (iia) are evolutionary graph theory (Lieberman et al 2005, Ohtsuki et al 2006), islands of equal size (Rousset and Billiard 2000) and certain models of group selection (Traulsen and Nowak 2006, Lehmann et al 2007b). Of course, the well-mixed population also fulfills (iia). Examples of population structures that fulfill (iib) are islands of variable size with global updating (Antal et al 2009, Taylor and Grafen 2010).

If (iia) holds, then $\partial w_\bullet/\partial s_j$ is independent of the state for all j and (21) becomes

$$\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} \langle s_\bullet s_j\rangle_0 > 0 \tag{22}$$

This is not yet in the form of (19) because in (22) we have identity by descent, Q_j, instead of the relatedness, R_j. However, if assumption (i) is fulfilled, the above condition is equivalent to (18) as follows

$$\begin{aligned}\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} R_j &= \sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} \frac{Q_j-\overline{Q}}{1-\overline{Q}} \\ &= \frac{1}{1-\overline{Q}}\left(\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} Q_j - \overline{Q}\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0}\right) \\ &= \frac{1}{1-\overline{Q}}\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} Q_j\end{aligned} \tag{23}$$

The last equality holds because in an additive game (where assumption (i) is fulfilled) $\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\Big|_{\delta=0} = 0$. This is the sum of all the effects of everyone in the population (including himself) on the actor, given that everyone is a cooperator. But if everyone is a cooperator the fitness of any individual is precisely $w_i = 1$, and hence its derivative with respect to δ is zero.

If (iib) holds (and again we stress the importance of global updating) then (21) becomes

$$\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} \langle s_\bullet s_j \mid i \text{ and } j \text{ interact } \rangle_0 > 0 \tag{24}$$

What we obtain is a local type of relatedness as also pointed out in the case of islands by Rousset and Billiard (2000). Thus (19) is generalized as

$$\sum_j \frac{\partial}{\partial \delta} \frac{\partial w_\bullet}{\partial s_j}\bigg|_{\delta=0} \tilde{R}_j > 0. \tag{25}$$

Here $\tilde{R}_j$ is the probability that the actor and individual j are identical by descent provided that they interact (they are on the same island, share the same tag, etc).

We have shown that the weak selection and low mutation limit of our general approach can also be calculated in terms of inclusive fitness if assumptions (i) and (ii) hold. Remember that as specified before, inclusive fitness theory cannot even be defined for non-vanishing selection; thus the assumption of weak selection is automatic. However, our weak selection result (14) holds for any mutation rate. Therefore, when assumptions (i) and (ii) are fulfilled, we can easily generalize the inclusive fitness condition (18) to any mutation and obtain

$$W_{IF} > uB_{IF}. \tag{26}$$

This can be interpreted as: the inclusive fitness effect has to be greater than the inclusive birth effect lost by mutation. We will expand on this in a forthcoming paper.

4 Example: a one dimensional spatial model

In this section we give a simple example of a model that satisfies assumptions (i) and (ii) and can thus be interpreted from both the natural selection and the inclusive fitness theory perspectives. We consider a population of N individuals on a one dimensional spatial structure. Each individual has two neighbors. To avoid boundary effects we connect the two end points to form a cycle. So far we have done the derivation for two general strategies A and B. Here however, we only need to consider the simplified Prisoner's Dilemma to make our point. Individuals are either cooperators, C or defectors, D. Cooperators pay a cost, c, for their neighbor to receive a benefit, b. Defectors pay no cost and distribute no benefit. We use death-birth (DB) updating: each time step an individual is picked at random to die; then the two neighbors compete proportional to their payoff to fill the empty spot with an offspring (Ohtsuki and Nowak 2006, Grafen 2007a).

Since death occurs at random, the death rate is $d_i = 1/N$ for all i. The birth rate is proportional to payoff. Individual i reproduces if one of his neighbors is picked to die and if he wins the competition for reproduction. We can write the birth rate of individual i as

$$b_i = \frac{1}{N}\left(\frac{f_i}{f_i + f_{i-2}} + \frac{f_i}{f_i + f_{i+2}}\right) \tag{27}$$

As before, the expected number of offspring of individual i is $w_i = 1 - d_i + b_i$. Let us now write the effective payoff of individual i

$$f_i = 1 + \delta(-2cs_i + bs_{i-1} + bs_{i+1}) \tag{28}$$

This effective payoff is the same for both approaches to the limit of weak selection discussed in Section 2. In the limit of weak selection, we can write the fitness of individual i as

$$w_i = 1 + \frac{\delta}{4N}\left(-4cs_i - bs_{i-3} + 2cs_{i-2} + bs_{i-1} + bs_{i+1} + 2cs_{i+2} - bs_{i+3}\right) \tag{29}$$

Since this is a process with constant death rate which moreover at neutrality has $b_i = d_i = 1/N$, we know from (15) that the condition that cooperation is favored on average, for weak selection, is equivalent to

$$\left\langle \sum_i s_i \frac{\partial w_i}{\partial \delta}\bigg|_{\delta=0} \right\rangle_0 = \left\langle \sum_i s_i(-4cs_i - bs_{i-3} + 2cs_{i-2} + bs_{i-1} + bs_{i+1} + 2cs_{i+2} - bs_{i+3}) \right\rangle_0 > 0 \quad (30)$$

This can be rewritten as

$$\begin{aligned} -4c\Big\langle \sum_i s_i^2 \Big\rangle_0 - b\Big\langle \sum_i s_i s_{i-3} \Big\rangle_0 + 2c\Big\langle \sum_i s_i s_{i-2} \Big\rangle_0 + b\Big\langle \sum_i s_i s_{i-1} \Big\rangle_0 \\ - b\Big\langle \sum_i s_i s_{i+3} \Big\rangle_0 + 2c\Big\langle \sum_i s_i s_{i+2} \Big\rangle_0 + b\Big\langle \sum_i s_i s_{i+1} \Big\rangle_0 > 0 \end{aligned} \quad (31)$$

These averages can be reinterpreted as probabilities as follows

$$\begin{aligned} \Big\langle \sum_i s_i^2 \Big\rangle_0 &= \frac{N}{2} \\ \Big\langle \sum_i s_i s_j \Big\rangle_0 &= \frac{N}{2}\Pr(s_i = s_j) \end{aligned} \quad (32)$$

The first identity holds because at neutrality the average number of cooperators equals the average number of defectors. The second identity consists of expressing the average in terms of a probability. Hence (31) becomes (after simplifying an $N/2$)

$$\begin{aligned} -4c - b\Pr(s_i = s_{i-3}) + 2c\Pr(s_i = s_{i-2}) + b\Pr(s_i = s_{i-1}) - \\ - b\Pr(s_i = s_{i+3}) + 2c\Pr(s_i = s_{i+2}) + b\Pr(s_i = s_{i+1}) > 0 \end{aligned} \quad (33)$$

Notice that so far this result holds for any mutation rate. This is what we would calculate on the cycle. What we are left to calculate are probabilities that individuals on the cycle have the same strategy. Notice that although this is an evolutionary process and two individuals in the stationary distribution share a common ancestor with probability 1, because of mutation, their strategies could have changed several times since they shared the common ancestor.

If mutation is very small ($u \to 0$) then the probability that two individuals have the same strategy in the stationary distribution is the same as the probability Q_{ij} that they are identical by descent (i.e. that they came from a common ancestor and have not mutated since). Hence the above probabilities, in the limit of low mutation, can be replaced by the respective Q_{ij}. However, as we explained before, for additive games, the Q_{ij} in (33) can be replaced by relatedness $R_{ij} = (Q_{ij} - \overline{Q})/(1 - \overline{Q})$ because $\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_i}{\partial s_j}\Big|_{\delta=0} = 0$. It is easy to test this for our particular cycle model, using (29):

$$\sum_j \frac{\partial}{\partial \delta}\frac{\partial w_i}{\partial s_j}\bigg|_{\delta=0} = -4c - b + 2c + b - b + 2c + b = 0 \quad (34)$$

Thus, in the limit of low mutation, we can rewrite (33) as

$$-4c - bR_{i-3} + 2cR_{i-2} + bR_{i-1} - bR_{i+3} + 2cR_{i+2} + bR_{i+1} > 0 \quad (35)$$

Here i is chosen at random to be the focal individual and $R_j = R_{ij}$. This is precisely $W_{dir} = W_{IF} > 0$ obtained by applying (19) to (29). We have discussed this example to show how when assumptions (i) and (ii) are satisfied, in the limit of weak selection, the two approaches give the same result.

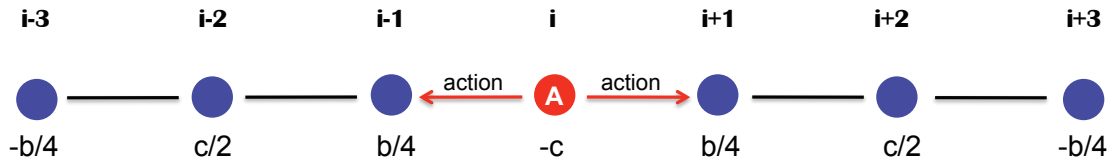

Figure 2: **Inclusive fitness is not easy to measure.** An empirical measurement of inclusive fitness has to contain every individual whose fitness is affected by the action (and not only those individuals whose payoffs are affected). Individual i is the actor A; individuals $i-1$ and $i+1$ are the direct recipients; individuals $i \pm 2$ and $i \pm 3$ do not interact with i but their fitness is affected by i's action due to indirect competition. For instance, i decreases his own fitness by $-c$; when it comes to compete for reproduction under DB updating, i competes with either $i-2$ or $i+2$ to fill the spot of $i \pm 1$ and hence a decrease in the fitness of i is an implicit benefit for both $i \pm 2$. Similarly, $i-1$ is the recipient of a benefit from i and he competes with $i-3$ for reproduction to fill the spot of $i-2$; hence, its benefit from i is detrimental to $i-3$ due to this competition. Taking into account all these effects one obtains the inclusive fitness expression (35).

Next we can express the condition (35) that cooperation is favored over defection as

$$\frac{b}{c} > \frac{4 - 2R_{i-2} - 2R_{i+2}}{R_{i-1} + R_{i+1} - R_{i-3} - R_{i+3}} \tag{36}$$

For symmetry reasons we have $R_{i-1} = R_{i+1}$, $R_{i-2} = R_{i+2}$ and $R_{i-3} = R_{i+3}$. Thus, we obtain

$$\frac{b}{c} > 2\frac{1 - R_{i-2}}{R_{i-1} - R_{i-3}} = \frac{2N-4}{N-4} \tag{37}$$

To obtain the final result we used the relatedness values as calculated by Grafen (2007a):

$$R_{i\pm3} = \frac{N^2 - 18N + 53}{N^2 - 1} \qquad R_{i\pm2} = \frac{N^2 - 12N + 23}{N^2 - 1} \qquad R_{i\pm1} = \frac{N-5}{N+1} \tag{38}$$

Thus we conclude that for DB updating on a cycle, cooperation can be favored provided that the benefit-to-cost ratio exceeds the threshold given by (37).

The same type of analysis can be performed for a Birth-Death (BD) updating on the cycle. In this case, an individual is picked to reproduce proportional to fitness and its offspring replaces one of the two neighbors at random. For this update rule however, there is no evolution of cooperation, despite the fact that the relatedness values are the same as before. This fact is also pointed out by Grafen (2007a).

The original analysis of Ohtsuki and Nowak (2006) is much simpler than both our analyses above. However, it is particular to low mutation on the cycle and not generalizable to more complex structures. Since the purpose of this paper is to discuss general approaches, we omit it here.

5 Hamilton's rule almost never holds

Below we use the same one dimensional spatial model to exemplify the fact that Hamilton's rule in the classical form, $bR > c$, almost never holds. Some inclusive fitness theoreticians seem to agree with this point and now propose that instead $W_{IF} > 0$ should be called Hamilton's rule (West and Gardner 2010). This section is not directed at theoreticians who now embrace $W_{IF} > 0$ as Hamilton's rule. Instead it is directed at empiricists that still try to test classical Hamilton's rule and at theoreticians who try to artifically reinterpret every result as classical Hamilton's rule.

Let us stay with our one dimensional spatial model. We find that both kin selection and natural selection give the same final result, because we have weak selection, an additive game and a 'special' population structure. For DB updating the condition that cooperators are favored over defectors can be written as (37)

$$\frac{b}{c} > 2\frac{1 - R_{i-2}}{R_{i-1} - R_{i-3}}$$

This condition is not as simple as Hamilton's rule, $bR > c$, but it is of the form $b \times (\text{something}) > c$. The only problem is that 'something' is not genetic relatedness, but a complicated function of relatedness. When these relatedness terms are calculated, the final result is, in the limit of large population size, $b/c > 2$. It is wrong, however, to think that relatedness on the cycle is $1/2$. That this is not the case can be seen by looking at the terms R_j in (38). The factor $1/2$ comes from evaluating the complicated function of relatedness given by (37).

If however we would not analyze the model of interactions on a one-dimensional structure, but instead we would wrongly think that Hamilton's rule holds, we would proceed as follows. We would calculate relatedness as is usually done: pick two individuals that interact and then compare their relatedness to the average relatedness in the population. This is precisely $R = R_{i-1} = (N-5)/(N+1)$. Then we would conclude that the condition that needs to be fulfilled for cooperation to prevail is $bR > c$ which leads to $\frac{b}{c} > (N+1)/(N-5)$. This however is wrong; the correct result is given by (37).

This situation is not particular to the cycle. In fact there are only very few, especially simple models, that have the property that the final result has the form $bR > c$, where R is relatedness (Rousset and Billiard 2000, Taylor and Grafen 2010). But for most other models (Ohtsuki et al 2006, Grafen 2007a, Taylor et al 2007a, Tarnita et al 2009a, Traulsen and Nowak 2008, Lehmann et al 2007a,b), all that can be obtained is a condition of the form

$$b \times (\text{something}) > c \tag{39}$$

This latter condition is not Hamilton's rule. As shown in Tarnita et al (2009b), the reason for a condition of this form is that for the limit of weak selection the final condition must be linear in the payoff values. For spatial processes the 'something' in (39) reflects the positive assortment created by a given model between individuals with the same strategy (Fletcher and Doebeli 2009, Nathanson et al 2009, Nowak et al 2010). Moving away from the limit of weak selection typically leads to conditions that are nonlinear in the payoff values.

Inclusive fitness theoreticians have also realized that Hamilton's rule does not hold in general and they caution against using it "naively", which would lead to mistakes (Roze and Rousset

2004). Gardner et al (2007), citing Taylor and Frank (1996) and Frank (1998), suggest that one should instead "use standard population genetics, game theory, or other methodologies to derive a condition for when the social trait of interest is favored by selection and then use Hamilton's rule as an aid for conceptualizing this result". We appreciate the proposal to simply use game theoretic/population genetics models based on natural selection, but we disagree that a forced reinterpretation of these results in terms of an artificially constructed variant of "Hamilton's rule" will help with any conceptualization. By artificially constructed variant we mean the following: when realizing that the usual $bR > c$ rule does not hold for a given model, Gardner et al (2007) propose that a modified rule $BR > C$ in fact holds, where R is the usual relatedness but B and C are the 'effective' costs and benefits calculated using statistical methods (which are not only unnecessary but also out of place in the analysis of a purely mathematical model). This method does not always work (we have not seen such a proposal for the cycle). Moreover, these effective costs and benefits unfortunately are very confusing and are typically functions of not only b and c but also of the relatedness R. Hence Hamilton's rule becomes $B(R)R > C(R)$, which makes it very complicated to separate any effects and it generally provides no intuition whatsoever. We argue that a simple but precise model with a careful natural selection-based analysis will suffice to provide any necessary conceptualization.

6 Relatedness measurements alone are inconclusive

Empirical biologists often seem to interpret inclusive fitness theory as suggesting that all that needs to be done is measure genetic relatedness and conclusive insights will emerge. Here we give a simple thought experiment to show that is not the case. An empirical measurement of relatedness in the absence of an understanding of the population dynamics can be misleading.

Consider three populations. Population 1 is well-mixed; any two individuals interact equally likely. Populations 2 and 3 are on a (one dimensional) spatial structure. Population 2 has birth-death (BD) updating: individuals reproduce proportional to payoff and the offspring replace randomly chosen neighbors. Population 3 has death-birth (DB) updating: random individuals die and the neighbors compete for the empty site proportional to their payoff. The empiricist measures relatedness in all three populations, but has no other information about the population dynamics.

The empiricist notes that the average relatedness of interacting individuals in population 1 is low and concludes that there is no scope there for evolution of cooperation. But for populations 2 and 3 the empiricist measures exactly the same high relative relatedness (Grafen 2007a) and hence concludes that cooperation is favored over defection in both cases. This is not true.

As we discussed in Section 4, cooperation can evolve in population 3 but not in population 2, although they generate the same measurements of relatedness (Ohtsuki and Nowak 2006, Grafen 2007a). Hence, the empirical measurement of relatedness without an actual knowledge of the underlying population dynamics can be very misleading.

7 When inclusive fitness fails

In this section we show that standard relatedness does not appear in general models, which do not fulfill the restrictive assumptions discussed above. Ultimately we recognize that ingenious theoreticians working in the area of kin selection might try to redefine 'relatedness' in new ways that allow them to see the cases below as 'obvious' inclusive fitness models. However, here we want

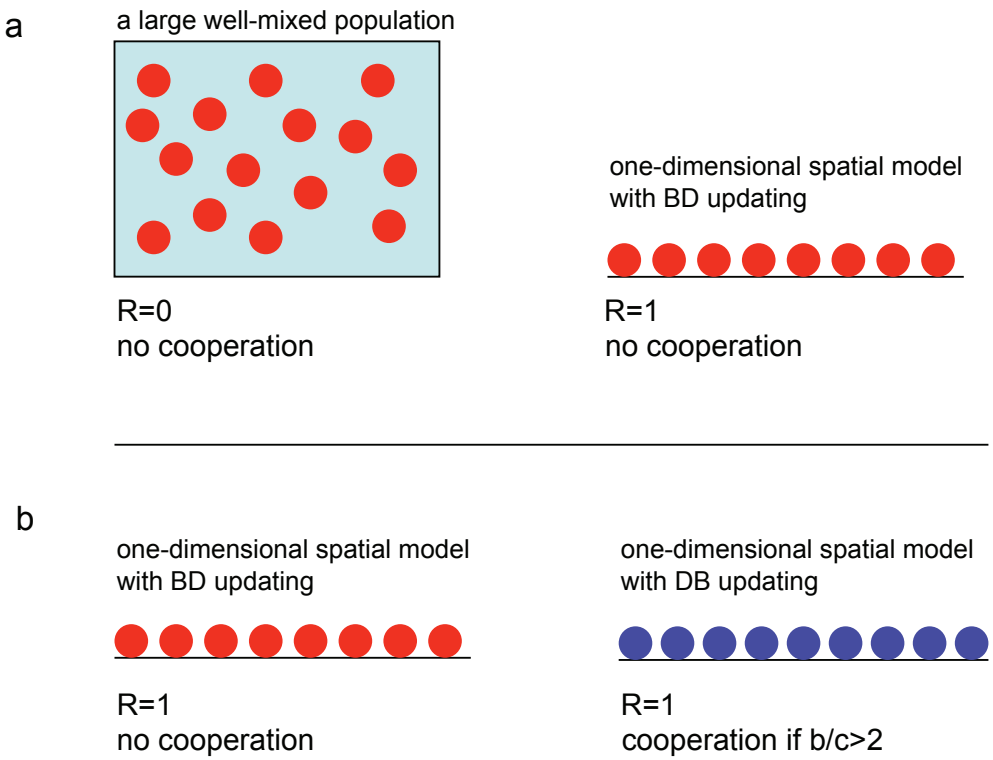

Figure 3: **Relatedness does not measure the ability of a population to support evolution of cooperation.** (a) A large well-mixed population has very low relatedness, $R = 0$, while a population that occupies a one dimensional spatial grid has maximum relatedness, $R = 1$. Nevertheless for birth-death (BD) updating both population structures are equally unable to support evolution of cooperation. (b) Now we compare two populations that are both arranged on a one dimensional spatial grid, and hence both populations have maximum relatedness, $R = 1$. But the first one uses birth-death (BD) updating and does not support evolution of cooperation, while the second one used death-birth (DB) updating and does support evolution of cooperation provided $b/c > 2$. BD updating means that individuals reproduce proportional to payoff and the offspring replace randomly chosen neighbors. DB updating means that individuals die at random and then the neighbors compete for the empty site proportional to payoff. Relatedness is $R = (Q - \bar{Q})/(1 - \bar{Q})$, where Q is the average relatedness of two individuals who interact and $\bar{Q}$ is the average relatedness in the population. These mathematical examples are chosen to be as simple as possible, but they make the more general point that relatedness data in the absence of a precise understanding of population dynamics are not very useful.

to show that there are clear limitations to inclusive fitness theory, if the R_j in (18) should still resemble a meaningful relatedness. If one allows for a definition of R_j that is not even remotely close to a relatedness that could be measured, then we do not see what is gained from a kin selection perspective. Pushing for such generalizations and extensions of inclusive fitness theory is not only cumbersome and confusing but ultimately useless for two reasons: theoretically they bring nothing new or even different from what is obtained with a simple, general, common sense result of the form (15) and empirically they have no value, because they do not use quantities that an empirical biologist could measure or call 'relatedness'.

7.1 Non-vanishing selection

Inclusive fitness theory requires weak selection for two reasons. First of all, even for pairwise interactions, in order to remove synergistic effects, one needs games with equal gains from switching. The limit of weak selection as described in Section 2.1 ensures that the games have this property. However, even if theoreticians restrict themselves to games with equal gains from switching, in a stochastic process, they still need to take the limit of weak selection. This is because the stochasticity introduces effects of competition between individuals and these are not independent of the 'helping' events and thus they cannot simply be added or subtracted. To make these events independent, one needs to be in the limit of weak selection.

Let us exemplify this second problem that arises for strong selection using the same one dimensional spatial model of Section 4. The birth rate of an individual is given by (27). In the limit of weak selection, the effective payoff (score) of individual i is given by (28). Both approaches agree with this in the limit of weak selection. However, a stronger selection variant has not yet been proposed and so we will not explicitly write it here but simply say that the effective payoff (score) of individual i is a function of his own strategy as well as those of the neighbors $f_i(s_{i-1}, s_i, s_{i+1})$. Let us consider DB updating. In this case the death rate is constant, $d = 1/N$. Then the average number of offspring of individual i is given by (29). In this case, even in a given state, one cannot write additively the effects of an individual on everyone else in the population, simply because when we take the derivative of w_i with respect to s_k, for some k, it is not necessarily independent of s_j for $j = 1, \ldots, N$. Let us exemplify this for simplicity, by taking the effects of an individual on himself

$$\frac{\partial w_i}{\partial s_i} = \frac{1}{N}\left(\frac{\frac{\partial f_i}{\partial s_i}(f_i + f_{i-2}) - \frac{\partial (f_i + f_{i-2})}{\partial s_i} f_i}{(f_i + f_{i-2})^2} + \frac{\frac{\partial f_i}{\partial s_i}(f_i + f_{i+2}) - \frac{\partial (f_i + f_{i+2})}{\partial s_i} f_i}{(f_i + f_{i+2})^2}\right) \tag{40}$$

Clearly this quantity still depends (in a highly non-trivial way) on some $s_i, s_{i\pm1}$ and $s_{i\pm2}$. Thus the effects of individuals on fitness are not independent (and certainly not linear) unless we are in the limit of weak selection. Therefore, non-vanishing selection is not just harder to calculate but it fails to lend itself to an inclusive fitness interpretation. Not only does the final result look more complicated than $bR > c$, but the concept of identity by descent does not arise in the calculation.

Moreover, we point out that very interesting results have already been obtained for any intensity of selection using the common sense approach based on natural selection (Ohtsuki and Nowak 2006, Traulsen et al 2008, Antal et al 2009b).

7.2 Non-additive games

By a non-additive game we mean one where interactions are either synergistic or are not necessarily pairwise. In an ant colony it is hard to imagine that pairwise interactions are sufficient to specify all fitness considerations. In fact, we expect that a synchronization of workers of different castes is necessary for success. This scenario cannot be covered by inclusive fitness theory.

Attention to such synergistic games was drawn by Queller (1985) who looked at 2-player games that do not have equal gains from switching. This is also discussed by Traulsen (2010). The idea was however expanded to non-additive, multiple player games by van Veelen (2009) who referred to them as 3-person stag hunt games. Gokhale and Traulsen (2010) also analyze multiple player one-shot games and point out the complex situations that can arise when pairwise interactions are insufficient to describe the dynamics.

The idea behind van Veelen's proposal is that it does not matter whether one or two of the players in a rock band rehearse – the band will sound lousy unless all three rehearse. A similar metaphor can be imagined for ants. Consider the following game. Imagine there are groups of size 3, all trying to accomplish a certain task, which can only be accomplished if all three individuals cooperate. A member of a group has the option to either cooperate and thus incur a cost c or defect. If all 3 individuals cooperate, they all incur the cost c but the task is accomplished and they are all rewarded with a benefit b. Thus the payoff of a CCC group is $(b-c, b-c, b-c)$. If neither cooperates, then they don't pay any cost but they also don't get any benefit. Then the payoff of a DDD group is $(0,0,0)$. However, and here comes the non-additivity of this game, if only one or two of the members cooperate and the others do not, the group does not accomplish the goal and so no one gets a benefit. The payoff of a CCD group is $(-c, -c, 0)$ and that of a CDD group is $(-c, 0, 0)$.[8]

There are n such groups; hence the total population size is $3n$. We label the individuals in a group by 1, 2 and 3 and we specify individual 1 in group k by s_1^k. There is no difference between the 1, 2 or 3 roles. Thus, the effective payoff of individual 1 in triple k is given by

$$f_1^k = 1 + \delta(-cs_1^k + bs_1^k s_2^k s_3^k) \tag{41}$$

Therefore a benefit arises only if all three members of the group cooperate. Clearly in such a game one cannot separate the effects of the action of the second player on the first from the effects of the action of the first on himself.

To make this point even more clear, let us consider the following dynamics. At each update step, an individual is picked to change his strategy. The way he changes the strategy is by choosing someone proportional to payoff and imitating their strategy. Thus, the death rate is constant $d = 1/3n$ and the birth rate of individual 1 in group k is $b_1^k = f_1^k/F$ where $F = \sum_k (f_1^k + f_2^k + f_3^k)$. Then the average number of offspring of individual 1 in pair k is

$$w_1^k = 1 - \frac{1}{3n} + \frac{f_1^k}{F} \tag{42}$$

which for weak selection becomes

$$w_1^k = 1 + \frac{\delta}{3n}\Big(- cs_1^k + bs_1^k s_2^k s_3^k - \frac{1}{3n}\sum_{j=1}^{n}\big[- c(s_1^j + s_2^j + s_3^j) + 3bs_1^j s_2^j s_3^j\big]\Big) \tag{43}$$

[8] As opposed to $(b-2c, b-2c, 2b)$ and respectively $(-2c, b, b)$ which would be the case in an additive game.

Then, for example, the effect of individual 1 in group $j \neq k$ on individual 1 in group k is

$$\frac{\partial}{\partial \delta}\frac{\partial w_1^k}{\partial s_1^j} \sim \frac{c}{9n^2} - \frac{b}{3n} s_2^j s_3^j \tag{44}$$

This shows that the effect of individual s_1^j on s_1^k is not independent of the action of other players; in particular, it depends on the simultaneous action of the other two individuals from group j. In this case, one cannot separate the effects of the action of one actor on all recipients and hence the type of inclusive fitness reasoning cannot be applied anymore. However, the general natural selection argument can be successfully applied.

Here we want to point out that one could extend the inclusive fitness definition to also include 'relatedness' of 3, 4, 5, ... individuals. However, the new definition will be far less intuitive; moreover, before applying such a definition one will need to know the actual model. Otherwise one can never know whether they need to consider relatedness of 3 individuals or of 4 or of 5 and so on. Such an analysis will end up being the same as the game theoretic analysis.

7.3 Generic population structure

If the population structure is not fixed, but dynamical, varying from one state to the other, and is more complex than islands, then one cannot separate relatedness from the structure. Queller (1994) already pointed out that for limited dispersal models, the concept of relatedness has to be local. Later Rousset and Billiard (2000) formalized this idea and showed that when the population is subdivided into islands (groups) one needs to compare the relatedness within a group to the relatedness in the overall population. However, this approach does not completely solve the problem. If individuals are always just on islands or on unweighted graphs (even if they are dynamical), it suffices to compare the relatedness of two people who interact to the average relatedness in the population. But if the population is on a dynamical, weighted graph, the situation becomes increasingly more complicated. Then it is not so easy to pick two individuals who interact and compare their relatedness to the average relatedness in the population, simply because two individuals might interact with varying weights (some individuals might have weight 1, others weight 1.5 and yet others weight 10). Since the structure is not fixed and these weights vary dynamically from one generation to the next, one cannot take a pair that has average weight either (because the average weight varies from one state to the next). Instead of needing to calculate quantities like

$$\text{Pr(two individuals who interact are identical by descent)} \tag{45}$$

one now has to calculate quantities like

$$\langle \text{Pr(that they are identical in state)} \times \text{(their average weight of interaction in that state)} \rangle_0 \tag{46}$$

averaged over all states of the system, at neutrality. Since the average interaction weight of two random individuals varies with state, the above average cannot be broken down into the product of the probability that they are identical by descent, times the average weight of interaction.

Let us consider a specific example. Tarnita et al (2009a) propose a model based on set memberships. We present here a somewhat simplified version of this model. Let us assume that individuals have exactly 2 tags [9]. There are M possible tags. Two individuals interact depending how many

[9] If they could only have one tag each, this would be an island model as described by Antal et al (2009a) and Taylor and Grafen (2010)

tags they have in common. If they share 0, 1 or 2 tags, they have interaction weight 0, 1, or 2, respectively. At each time step, a randomly chosen individual updates his strategy and tags. He imitates someone in the population proportional to payoff to obtain new tags and a new strategy. Therefore the population structure is dynamical. Depending on the state all individuals could have the same two tags or they might be spread out over many different tags.

The interactions between individuals are dynamical, in the sense that they change from a state to another. Individuals who might interact twice this week, might not interact at all next week. In each state we specify by $v_{ij} \in \{0, 1, 2\}$ the interaction weight of i and j. For simplicity, let us assume that $v_{ij} = v_{ji}$. If $v_{ij} = 0$ then i and j do not interact. These interaction weights are dynamical in the sense that they change as a consequence of evolutionary updating. Then, in a given state, the effective payoff (score, fecundity) of individual i is given by

$$f_i = 1 + \delta \sum_j v_{ij}(-cs_i + bs_j) \tag{47}$$

Since the death rate is at random, we have $d_i = 1/N$. The birth rate is proportional to effective payoff and all individuals compete for reproduction. Therefore, we have

$$b_i = \frac{f_i}{F} \tag{48}$$

Here F is the total payoff in that given state. The average fitness of individual i is given by

$$w_i = 1 - \frac{1}{N} + \frac{f_i}{F} \tag{49}$$

For the limit of weak selection we obtain

$$\begin{aligned} w_i &= 1 + \frac{\delta}{N}\Big[\sum_j v_{ij}(-cs_i + bs_j) - \frac{1}{N}\sum_j \sum_k v_{jk}(-cs_j + bs_k)\Big] \\ &= 1 + \frac{\delta}{N}\Big[s_i\Big(-c + \frac{c}{N} - \frac{b}{N}\Big)\sum_j v_{ij} + \sum_{j\neq i} s_j\Big(bv_{ij} - \frac{b-c}{N}\sum_k v_{jk}\Big)\Big] \end{aligned} \tag{50}$$

Clearly here the action of individual j on individual i depends on the dynamical structure. The effect of j on i is

$$\frac{\partial w_i}{\partial s_j} = \begin{cases} \big(-c + \frac{c}{N} - \frac{b}{N}\big)\sum_k v_{ik} & \text{if } j = i \\ bv_{ij} - \frac{b-c}{N}\sum_k v_{jk} & \text{if } j \neq i \end{cases} = \begin{cases} (-c(N-1) - b)v_{ik} & \text{if } j = i \\ bv_{ij} - (b-c)v_{jk} & \text{if } j \neq i \end{cases} \tag{51}$$

The second equality comes from replacing $\sum_l v_{kl} = N v_{kl}$; we do this by picking a random individual instead of summing over all of them. Thus, the effect of the actor on himself is proportional to how much he interacts with a random individual; the effect of a random individual on the actor depends on how much he interacts with the actor and on how much he interacts with someone at random.

Since this is a process with constant death, in the limit of weak selection but for any mutation, the condition for cooperators to be favored over defectors is given by (15). The individuals are symmetric so instead of summing over all individuals as in (15), we can pick a random representative

one, which we denote by •. Using (50) together with (51) we obtain the condition for cooperators to be favored over defectors:

$$(-c(N-1)-b)\langle v_{\bullet k}s_{\bullet}\rangle_0 + \sum_{j\neq\bullet}\left\langle (bv_{\bullet j}-(b-c)v_{jk})s_{\bullet}s_j\right\rangle_0 > 0 \tag{52}$$

This expression does not look like $W_{IF} > 0$ anymore because what usually is identity by descent now depends on the structure, containing quantities of the form $\langle v_{\bullet j}s_{\bullet}s_j\rangle_0$ which can no longer be interpreted as local relatedness. This is because $v_{\bullet j}$ changes from each state to the next and to calculate such quantities, one needs to pick a random individual j in every state and multiply the number of tags the actor and j have in common with the probability that they are identical in state. This quantity is then averaged over all possible states.

In some sense, what a dynamical structure captures is the fact that I may be very related to my sister but if I see her once a year now and maybe 100 times next year, this can make a huge difference overall. Thus, for such dynamical structures genetic relatedness alone is not important but it somehow has to be weighted by the varying intensities of interaction.

8 Group selection is not kin selection

Group selection arises whenever there is competition not only between individuals but also between groups. Group selection is part of the more general concept of multi-level selection. There has been a long and ongoing debate between scientists who work on group selection and kin selection (Wynne-Edwards 1962, Wilson 1975, Killingback et al 2006, Grafen 2007, Traulsen and Nowak 2006, Lehmann et al 2007b, Wilson and Wilson 2007, Bijma and Wade 2008, Goodnight et al 2008, West et al 2008, West et al 2009, Wild et al 2009, van Veelen 2009, Traulsen 2010, Wade et al 2010) with the kin selection side claiming that group selection and kin selection are identical approaches. In the light of what we have shown here, we hope to settle this debate. Group selection models, if correctly formulated, can be useful approaches to studying evolution. Moreover, the claim that group selection is kin selection is certainly wrong.

Group selection models can be formulated for any intensity of selection (Traulsen et al 2008). Since, when dealing with stochastic processes, inclusive fitness theory works only for the limit of weak selection, results for groups with non-vanishing selection cannot be replicated by such a theory. As we have pointed out weak selection results are interesting, but they do not offer a complete picture of evolution. Hence, from the start, kin selection cannot possibly cover the same ground as group selection.[10]

If however we limit ourselves to weak selection, group selection can be interpreted in terms of inclusive fitness calculations only if assumptions (i) and (ii) hold. But it is very easy and very natural to formulate group selection models that violate the assumptions needed for an inclusive fitness calculation. Group selection models could contain non-additive games or dynamical population structures, where individuals interact with different intensities. Thus, even for weak selection,

[10]It is worth noting that for deterministic, replicator equation-type models, van Veelen (2009) shows that group selection models and inclusive fitness models give the same prediction even for non-vanishing selection, as long as the inclusive fitness theory restricts itself to the non-generic case of games with equal gains from switching. However, such a result does not hold for stochastic models, where the inclusive fitness method requires weak selection not only to obtain equal gains from switching, but also to make the interactions and competitions between individuals independent and additive (Section 7.1).

group selection cannot in general be described by inclusive fitness calculations. The claim that they are identical, which has often been made (Lehmann et al 2007b, West et al 2008, Wild et al 2009), is wrong.

9 Summary

We emphasize the following points:

- **Inclusive fitness is just another method of accounting.** The fact that an inclusive fitness calculation works for a particular model does not necessarily imply that 'kin selection is at work'. Inclusive fitness theoreticians have been inconsistent about this point. Originally, they suggested that it is just another method of calculation and stressed that Hamilton's 1964 paper is 'devoted to proving that the alternative accounting procedure that underlies inclusive fitness gives the same answer as the standard and logically prior procedure' (Grafen 1984). We agree with this perspective but add that, as we have proved in this paper, the inclusive fitness method cannot be used as widely as the logically prior and more general procedure based on natural selection.

  Of course, theoreticians are free to use any method of calculation as long as they employ it correctly and do not make unjustified statements claiming a 'general principle' for the evolution of cooperation (Lehman et al 2007a,b, Wild et al 2009, West et al 2008, Gardner 2009, West and Gardner 2010). A method of calculation which is arguably more cumbersome and confusing is not a general principle, much like the ptolemaic epicycles in the solar system were not a general principle either and became superfluous under Newtonian mechanics.

  We have a similar situation in this debate. The epicycles of inclusive fitness calculations are not needed, given that we can formulate precise descriptions of how natural selection acts in structured populations.

- **Inclusive fitness is not nearly as general as the game theoretic approach based on natural selection.** As we have pointed out here, the concept of inclusive fitness only leads to correct results if a number of constraining assumptions hold. These are the limit of weak selection together with assumptions (i) additive games and (ii) very simplistic population structures.

- **Inclusive fitness is often wrongly defined.** Inclusive fitness is NOT the sum of an individual's offspring plus the offspring of the relatives. It only includes an individual's own offspring that come as a result of his own actions, but not as a result of the help received from others. Thus my inclusive fitness is:

  (my offspring resulting from my own actions but NOT from the help I receive from others)
  $+ R \times$ (my relatives' offspring resulting from my helping them)

  This definition is somewhat artificial and leads to significant confusion in practice. Grafen (1984) warned empiricists and theoreticians against incorrect usage of the term and advised empiricists against using inclusive fitness altogether. Instead, they are told to use Hamilton's rule. This suggestion, however, brings us to another problem.

- **Hamilton's rule almost never holds.** For the limit of weak selection, we often find that cooperation is favored over defection provided 'something' is greater than the cost to benefit ratio. This result is not the consequence of inclusive fitness theory, but arises because of the linearity introduced by weak selection (Wild and Traulsen 2007, Tarnita et al 2009b). As we have shown here, however, 'something' is almost never 'relatedness', even if we are in a special situation where inclusive fitness can be formulated.

  The validity of Hamilton's rule is also challenged by economists. Alger and Weibull (2009) show that the evolutionarily stable degree of altruism (threshold cost-to-benefit ratio) is lower than the degree of relationship; moreover, it strongly depends on the environment, which is not something inclusive fitness theoreticians consider. The analysis of Alger and Weibull (2009) suggests a reason why weaker family ties may have developed in harsher climates, and how this may have induced stronger economic growth, according to Macfarlane (1978, 1992).

- **Relatedness measurements without a model of population dynamics are inconclusive.** While relatedness measurements in the field can provide useful information about population structure, they do not provide immediate information for the evolution of cooperation. Relatedness measurements always must be interpreted in the context of a model of evolutionary dynamics. Otherwise they can lead to meaningless conclusions. Relatedness measurements alone are not a test for inclusive fitness theory.

- **Inadequacy of inclusive fitness.** Let us consider a situation where one individual pays a cost to benefit another individual. If the two individuals are related, then there is a probability that they might both carry the same gene that affects altruistic behavior, and hence that gene could have a fitness advantage. We have shown that inclusive fitness theory cannot decide in general if an allele that makes you help a relative is favored by natural selection or not. Instead we need a calculation that is based on a precise description of population structure and dynamics.

Part B – Empirical tests reexamined

Faith in the central role of kinship in social evolution, defined in one manner or other, has led to the reversal of the usual order in which biological research is conducted. The proven best way in evolutionary biology is to define a problem arising during empirical research, then select or devise the theory that is needed to solve it. Almost all research in inclusive fitness theory has been the opposite: hypothesize the key roles of kinship and kin selection, then look for evidence to test that hypothesis.

The most basic flaw in this approach is failure to consider multiple competing hypotheses. Often the measurement of pedigree kinship becomes a surrogate for an in-depth study of natural history of the species. When the data do not fit, elaborations of inclusive-fitness theory can be constructed that make them fit. The results of the elaborations are a ptolemaic theory, constructed of epicycles to keep relatedness at the center of evolving social systems. With enough of epicycles, they can even be made to fit the data, but, as in the geocentric theory of the cosmos, at the cost of using the wrong starting assumption. This misstep in logic is known as 'Affirming the Consequent. When biological details of particular cases are examined before inclusive fitness theory is applied, alternative explanations from standard natural selection theory come quickly to attention. We have examined an array of the most meticulously analyzed cases presented by various authors as evidence for the success of kin selection theory. Without exception it has been easy to find weaknesses in each case, suggesting that analyses of empirical observations based on kin selection theory alone are typically inconclusive. We know of no case that presents compelling evidence for the explanatory adequacy of kin selection and inclusive fitness theory. Three representative examples of the inadequacy are the following.

First, Hughes et al (2008) argue that the origin of eusociality is driven by close kinship, because in basal clades of eusocial ants, bees and wasps, queens mate only with a single male and therefore produce a colony of closely related individuals. The authors present their data as correlative evidence of kin selection promoting the evolution of eusociality. However, comparable data were not provided for solitary clades, including sister clades of the eusocial examples, hence there were no controls for the retrodiction of kin selection. In fact, it is logical to suppose that such queens also mate with one male only, and for a reason unrelated to kin selection: prolonged mating excursions increase the risk to young females from predators. Thus even at best it is not proved that single mating and the resulting close relatedness of offspring is an important factor in promoting the evolution of eusociality. Of equal importance, Hughes et al (2008) point to the origin of multiple-male matings practiced by queens of many clades with advanced colonial organization. This, they conclude, indicates the relaxation of kin selection in later stages of evolution. But they overlook the much simpler explanation that in species with exceptionally large worker populations, queens need multiple matings in order to store enough sperm. Further, and oddly, many studies have found that, as a rule, members of social insect colonies cannot recognize their own degree of pedigree relatedness to their nestmates (Hölldobler & Wilson 2009, Ratnieks et al 2006). For ants at least, membership in a particular colony is determined by the overall colony odor learned by each adult during the first several days following her emergence from the pupa as an adult. For this reason it is possible to experimentally produce colonies whose members differ from one another radically, even belonging to different species. Such is the basis of slavery practiced by some parasitic ant species (Hölldobler & Wilson 2009, Ratnieks et al 2006). Is such learning just a proxy for staying close to genetic kin? Perhaps, but a more likely selection force is communal fidelity and defense of the nest, as we have documented in the main text.

In another, very different setting, a meticulous experimental analysis using the periodically subsocial eresid spider Stegodyphus lineatus, has demonstrated that groups of sibling spiderlings extract more nutrients from communal prey than do spiderlings of artificially mixed parentage (Schneider & Bilde 2008). Because injecting digestive enzymes is costly, the authors suggest that individuals withhold their enzymes to avoid exploitation by strangers. The authors accept the kin selection hypothesis. The problem is that in their natural habitat these spiderlings (necessarily coming from the same mother) never find themselves in mixed groups. Hence, we would not expect any specific adaptation to deal with such a situation. A much simpler explanation for the observed reduction in communal intake is given by discordance among unrelated individuals (meaning they are not working together efficiently). Such discordance is a general principle, which is often overlooked by kin selection based explanations. Genetically diverse groups (especially those that are generated artificially in an experimental situation) are prone to be less harmonious, because the individuals are not necessarily adapted to work with each other.

A third process that can lead to seeming kin-based altruism but is more simply and realistically explained otherwise is the expectation of inheritance. In a small percentage of bird and mammal species, offspring remain at the nest of their birth and assist their parents in rearing additional broods. They thereby delay reproduction on their own while increasing reproduction of their parents. In one interpretation, Griffin and West (2003) attribute the phenomenon to kin selection, and bolster their argument by demonstrating a positive correlation across species between closeness of kinship and the amount of help provided to parents by the stay-at-homes. However, more thorough, previously published studies, covering life history data in a wide range of species, had already arrived at a simpler explanation of why, under certain conditions unrelated to kin selection, the persistence of adult young at the natal nest is favored. The conditions include unusual scarcity either of nest site or territory or both, generally low adult mortality, and relatively unchanging conditions in a stable environment. After prolonged residence, the helpers inherit the nest or territory upon the death of the parents (Hatchwell & Komdeur 2000). The positive correlation across species between kinship and helping reported by Griffin and West, is based on a few widely scattered data points, but if upheld might well be explained by the common practice of the floater strategy in some species (Hatchwell & Komdeur 2000), in which individuals move about nests and spread the amount of help given.

Part C – A mathematical model for the origin of eusociality

We consider a species of solitary insects (wasps or others) where fertilized females build nests and raise their offspring by progressive provisioning. Once the larvae hatch, the offspring leave the nest. We call this the solitary life cycle. As outlined in the main text, we then assume that a mutant arises, where the offspring do not necessarily leave the nest. Those young that stay at the nest engage in the task of helping their mother to raise her subsequent offspring. Hence, they do not reproduce themselves, but sacrifice their reproductive potential to help another individual to reproduce. We call this the eusocial life cycle. We study the conditions for natural selection to favor the eusocial strategy over the solitary one.

At first we design a model with asexual reproduction in order to understand the essence of the problem. Subsequently we develop a model with sexual reproduction, taking into account the haplodiploid genetics of Hymenoptera. In both cases, we see that very different mathematical structures arise from those that were considered by kin selection theorists over the last decades. The kin selectionist's framing of the problem in terms of cost and benefit of the worker is not natural, and an inclusive fitness type calculation is not necessary.

Our model differs from previous approaches. Wade (1978) performs an interesting population genetic analysis of a situation where the broods of several females develop in close proximity and cooperate with each other to some extent. Craig (1979, 1983) explores parental manipulation and subfertility, which are potential mechanisms for the evolution of eusociality (Alexander 1974, Michener & Brothers 1974, West-Eberhard 1975, Charnov 1978) sometimes subsumed under kin selection (Bourke & Franks 1995). Gadagkar (1990) argues that eusociality could have arisen via several different mechanisms and what is important is not only its emergence but also its maintenance. Lehmann et al (2008) study a model where sterile workers migrate between colonies offering their help to different queens, but this is an unlikely biological scenario for the evolution of eusociality.

The theory that we develop here represents the simplest possible and most direct approach to evaluate how natural selection acts on alleles prescribing social behavior. The target of selection is neither the phenotypic trait of the queen in particular, nor that of the colony, but the collectivity of traits that modify social behavior at both these levels. The traits can be evaluated separately and together.

10 Asexual reproduction

10.1 A simple linear model

We consider deterministic evolutionary dynamics described by ordinary differential equations. Let x_0 denote the abundance of solitary females. They reproduce at rate b_0 and die at rate d_0. Let x_i denote the abundance of eusocial nests (colonies) of size i. Here $i = 1, 2, ...$ represents the number of individuals working at the colony including the queen. Thus, x_1 denotes nests with single eusocial queens, while x_2 denotes nests where a queen has one worker, and so on.

A eusocial queen in a nest of size i has reproductive rate b_i and death rate d_i. We assume that the presence of workers allows the queen to stay with the nest, which might increase her rate of oviposition and reduce her death rate. There is less risk of predation both for the queen and for the eggs, which can be guarded by the queen and other workers. The queen in particular can redirect

her resources from foraging to laying eggs and protecting them. The offspring of the eusocial queen stay with the nest with probability q and leave the nest with probability $1-q$. In the latter case they start their own nest.

Evolutionary dynamics of the two strategies are described by the following system of linear differential equations

$$\begin{aligned} \dot{x}_0 &= (b_0 - d_0)x_0 \\ \dot{x}_1 &= \sum_{i=1}^{\infty} b_i(1-q)x_i - b_1 q x_1 - d_1 x_1 \\ \dot{x}_i &= b_{i-1}qx_{i-1} - b_i q x_i - d_i x_i \qquad i = 2, 3, ... \end{aligned} \tag{53}$$

The strategy with the faster growth rate wins eventually. The exponential growth rate of the solitary strategy is $b_0 - d_0$, while that of the eusocial strategy is given by the largest eigenvalue, λ, of the matrix

$$M = \begin{pmatrix} b_1(1-q) - (b_1 q + d_1) & b_2(1-q) & b_3(1-q) & b_4(1-q) & \dots \\ b_1 q & -(b_2 q + d_2) & 0 & 0 & \dots \\ 0 & b_2 q & -(b_3 q + d_3) & 0 & \dots \\ 0 & 0 & b_3 q & -(b_4 q + d_4) & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{54}$$

If $\lambda > b_0 - d_0$ then eusociality wins over solitary. If $\lambda < b_0 - d_0$ then solitary wins. The case $\lambda = b_0 - d_0$ is ungeneric. Matrix, M, can be seen as a 'mutation-selection matrix' as it occurs for example in quasispecies theory (Eigen & Schuster 1977). Formally speaking the growth of colonies, described by the parameter q, denotes a 'mutation term' between colonies of different size. The largest eigenvalue of M is the fitness of the 'quasispecies'.

Whether or not eusociality is selected depends on how the demographic parameters of the queen change with colony size. One possibility is to consider a simple step function with a critical colony size, m. For small colonies, $i < m$, the key parameters of the eusocial queen are the same as those of solitary females: $b_i = b_0$ and $d_i = d_0$. For large colonies, $i \geq m$, the eusocial queen has an increased fecundity and a reduced death rate: $b_i = b > b_0$ and $d_i = d < d_0$.

A necessary condition for the evolution of eusociality is

$$b - k_m d > k_m(b_0 - d_0) \tag{55}$$

The number k_m depends on the critical colony size, m, where the advantages of eusociality arise. It turns out that k_m grows exponentially with m: we have $k_2 = 2$, $k_3 = 9$, $k_4 = 28$, $k_5 \approx 76$, $k_6 \approx 196$ and so on. It is essential that already small colonies increase the reproductive rate, b, of the queen. Otherwise it is very hard for eusociality to be selected. Despite its obvious and intuitive advantages it is in fact not easy to evolve eusociality.

Condition (55) is necessary but not sufficient. If (55) holds then there exist values of q that allow eusociality to win. There is a lower bound, q_{min}, and an upper bound, q_{max}. Eusociality wins if $q_{min} < q < q_{max}$. For $m = 2$ we find $q_{min} = 0$ and $q_{max} = 1 - 2(b_0 - d_0 + d)/b$. For $m \geq 3$ we have $0 < q_{min} < q_{max} < 1$. Depending on the value of b we observe that eusociality sometimes wins only for a narrow range of q values.

10.2 Adding density limitation

We can add density limitation by multiplying each birth term with a factor ϕ that represents declining food resources as the total population size increases. A natural possibility is $\phi = 1/(1 + \eta X)$ where η is a parameter that scales the size of the system and $X = x_0 + \sum_i i x_i$ is the total population size. The system with density limitation can be written as

$$\begin{aligned} \dot{x}_0 &= (b_0\phi - d_0)x_0 \\ \dot{x}_1 &= \sum_{i=1}^{\infty} b_i\phi(1-q)x_i - b_1\phi q x_1 - d_1 x_1 \\ \dot{x}_i &= b_{i-1}\phi q x_{i-1} - b_i\phi q x_i - d_i x_i \qquad i = 2, 3, ... \end{aligned} \tag{56}$$

The mathematical analysis is very similar to that in the previous section. There is no coexistence between the two strategies. One strategy will exclude the other. We must find the largest eigenvalue, λ, of the matrix

$$M = \begin{pmatrix} b_1(1-2q)/d_1 & b_2(1-q)/d_1 & b_3(1-q)/d_1 & b_4(1-q)/d_1 & \dots \\ b_1 q/d_2 & -b_2 q/d_2 & 0 & 0 & \dots \\ 0 & b_2 q/d_3 & -b_3 q/d_3 & 0 & \dots \\ 0 & 0 & b_3 q/d_4 & -b_4 q/d_4 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{57}$$

As before, eusociality is selected if $\lambda > b_0 - d_0$. Let us consider the same step function as before, where the advantages of eusociality arise if the colony size is at least m. We find a similar necessary condition for the evolution of eusociality, $b/d > k_m(b_0/d_0)$. The numbers k_m are the same as before. The expressions for q_{min} and q_{max} are also slightly different. For example, for $m = 2$ we have $q_{min} = 0$ and $q_{max} = 1 - 2b_0 d/(bd_0)$. Again, for $m \geq 3$ we find that eusociality wins only for a restricted range of q values, $0 < q_{min} < q < q_{max} < 1$.

10.3 Adding worker mortality

In the previous models the mortality of workers was somehow folded into the reproduction rate of the queen, but here we explicitly assume that workers die at rate α. We obtain

$$\begin{aligned} \dot{x}_0 &= (b_0\phi - d_0)x_0 \\ \dot{x}_1 &= \sum_{i=1}^{\infty} b_i\phi(1-q)x_i - b_1\phi q x_1 - d_1 x_1 + \alpha x_2 \\ \dot{x}_i &= b_{i-1}\phi q x_{i-1} - b_i\phi q x_i - d_i x_i - \alpha(i-1)x_i + \alpha i x_{i+1} \qquad i = 2, 3, ... \end{aligned} \tag{58}$$

As before we use $\phi = 1/(x_0 + \eta \sum_i i x_i)$. Again there is no coexistence between the solitary strategy and the eusocial one. There are two equilibria. At the solitary equilibrium x_0 is positive and all x_i (with $i = 1, 2, ...$) are zero. At the eusocial equilibrium x_0 is zero and all x_i (with $i = 1, 2, ...$) are positive. One equilibrium is stable and the other one unstable with respect to invasion by the opposite strategy. The stable equilibrium is the one that establishes the larger total population size. This criterium specifies whether or not eusociality is favored by natural selection.

Figure 4 shows how the winning strategy changes as a function of the parameter q, which denotes the probability that a eusocial offspring stays with the nest. For small values, $0 < q < 0.36$, and for large values, $0.9 < q < 1$, the solitary strategy wins. For the intermediate region, $0.36 < q < 0.9$, the eusocial strategy wins. In this numerical example the solitary female has reproductive rate $b_0 = 0.5$, which means she lays one (surviving) egg every other day. Her death rate is $d_0 = 0.1$, which implies an average life time of 10 days. The solitary female has to lay eggs and then go out to search for food, which exposes both the eggs and herself to risk from predation. The eusocial queen has the same parameters unless the colony reaches a critical size, here $m = 3$. Subsequently her reproductive rate increases 8-fold and her death rate decreases 10-fold. Thus for all $i \geq m$ we assume $b_i = 4$ and $d_i = 0.01$. She has less risk of predation and can devote her resources to laying eggs and protecting them. We assume that the workers in the eusocial colony have the same death rate as solitary individuals, $\alpha = 0.1$. Finally we assume, if the queen dies then the colony disappears.

Note that intermediate q values are needed for eusociality to evolve. The intuitive explanation is as follows. For low q there is only a small probability that the colony reaches the critical size, m, where the advantage of eusociality begins. On the other hand, if the value of q is too large then the colonies produce too few new queens. Of course, the disadvantage of eusociality is that some of the offspring (workers) do not reproduce; they are subject to worker mortality and they die when the queen dies. Therefore, intermediate values of q allow the evolution of eusociality.

The situation that we observe here is not a standard cooperative dilemma (Hauert et al 2006, Nowak 2006, Nowak et al 2010). An offspring that stays with the nest could be seen as a 'cooperator', an offspring that leaves could be seen as a 'defector', but the optimum strategy of eusociality has an intermediate level of cooperators and defectors. 'Defectors' (new queens) are needed for the reproduction of eusociality.

11 Sexual reproduction and haplodiploid genetics

We now develop a model that takes into account sexual reproduction and the haplodiploid genetics of Hymenoptera. We study competition between two alleles: the wildtype allele, A, and the mutant allele a. The mutant allele disrupts the dispersal behavior of females as described in the main text. It can either act in a dominant or recessive way. Denote by q_1, q_2, q_3 the probabilities that AA, Aa and aa females stay with the nest. If the mutation is dominant, we have $q_1 = 0$ and $q_2 = q_3 > 0$. If the mutation is recessive, we have $q_1 = q_2 = 0$ and $q_3 > 0$.

There are three types of females, AA, Aa, aa, and two types of males, A and a. There are six types of fertilized females (queens), which we denote by AA-A, AA-a, Aa-A, Aa-a, aa-A and aa-a. The first two letters specify her own genotype, while the third letter specifies the genotype of the sperm she has received and stored. For example, AA-a means that an AA female has mated with an a male. Such an AA-a queen produces Aa females and A males. In contrast an Aa-A queen produces equal proportions of AA and Aa females and equal proportions of A and a males.

Let us introduce the following notation: $X_{AA\text{-}A,i}$ denotes the abundance of colonies of size i founded by an AA-A queen; similarly $X_{AA\text{-}a,i}$ denotes the abundance of colonies of size i founded by an AA-a queen, and so on. The queen in a colony of size i has birth rate b_i and death rate d_i. The mortality of workers is given by α. The abundances of virgin queens (females who have left the nest and try to mate) is given by x_{AA}, x_{Aa} and x_{aa}. The abundance of males is given by y_A and y_a. The parameter β characterizes the rate of successful mating.

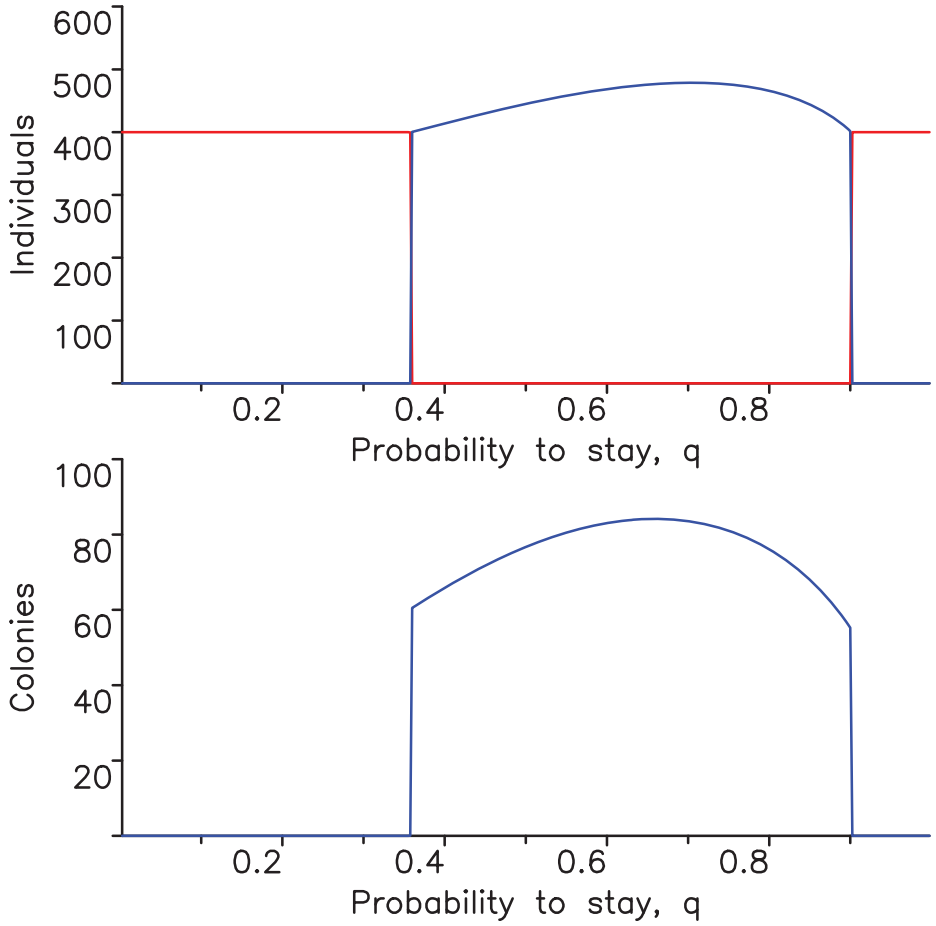

Figure 4: Eusociality evolves for intermediate values of q, which is the probability that offspring stay with the nest. Furthermore, the eusocial queen must have a dramatically increased rate of oviposition, b. The figure shows equilibrium values of the model given by (58). Parameter values are $b_0 = 0.5$, $d_0 = 0.1$, $m = 3$, $b = 4$, $d = 0.01$ and $\alpha = 0.1$. Hence, eusocial queens that have at least two workers have an 8-fold increased rate of laying eggs and a 10-fold reduced death rate. The upper panel shows the number of individuals, where red indicates solitary, x_0, and blue eusocial, $\sum_i ix_i$. The lower panel shows the total number of colonies with at least one worker, $\sum_i x_i$, where $i = 2, 3, \ldots$. Note that eusociality is selected if approximately $0.36 < q < 0.9$.

We have the following system (the index i runs from $2, ..\infty$):

$$
\begin{aligned}
\dot{X}_{AA\text{-}A,1} &= \beta x_{AA} y_A - b_1 q_1 X_{AA\text{-}A,1} - d_1 X_{AA\text{-}A,1} + \alpha X_{AA\text{-}A,2} \\
\dot{X}_{AA\text{-}A,i} &= q_1(b_{i-1} X_{AA\text{-}A,i-1} - b_i X_{AA\text{-}A,i}) - d_i X_{AA\text{-}A,i} - \alpha(i-1) X_{AA\text{-}A,i} + \alpha i X_{AA\text{-}A,i+1} \\
\dot{X}_{AA\text{-}a,1} &= \beta x_{AA} y_a - b_1 q_2 X_{AA\text{-}a,1} - d_1 X_{AA\text{-}a,1} + \alpha X_{AA\text{-}a,2} \\
\dot{X}_{AA\text{-}a,i} &= q_2(b_{i-1} X_{AA\text{-}a,i-1} - b_i X_{AA\text{-}a,i}) - d_i X_{AA\text{-}a,i} - \alpha(i-1) X_{AA\text{-}a,i} + \alpha i X_{AA\text{-}a,i+1} \\
\dot{X}_{Aa\text{-}A,1} &= \beta x_{Aa} y_A - b_1 \frac{q_1+q_2}{2} X_{Aa\text{-}A,1} - d_1 X_{Aa\text{-}A,1} + \alpha X_{Aa\text{-}A,2} \\
\dot{X}_{Aa\text{-}A,i} &= \frac{q_1+q_2}{2}(b_{i-1} X_{Aa\text{-}A,i-1} - b_i X_{Aa\text{-}A,i}) - d_i X_{Aa\text{-}A,i} - \alpha(i-1) X_{Aa\text{-}A,i} + \alpha i X_{Aa\text{-}A,i+1} \\
\dot{X}_{Aa\text{-}a,1} &= \beta x_{Aa} y_a - b_1 \frac{q_2+q_3}{2} X_{Aa\text{-}a,1} - d_1 X_{Aa\text{-}a,1} + \alpha X_{Aa\text{-}a,2} \\
\dot{X}_{Aa\text{-}a,i} &= \frac{q_2+q_3}{2}(b_{i-1} X_{Aa\text{-}a,i-1} - b_i X_{Aa\text{-}a,i}) - d_i X_{Aa\text{-}a,i} - \alpha(i-1) X_{Aa\text{-}a,i} + \alpha i X_{Aa\text{-}a,i+1} \\
\dot{X}_{aa\text{-}A,1} &= \beta x_{aa} y_A - b_1 q_2 X_{aa\text{-}A,1} - d_1 X_{aa\text{-}A,1} + \alpha X_{aa\text{-}A,2} \\
\dot{X}_{aa\text{-}A,i} &= q_2(b_{i-1} X_{aa\text{-}A,i-1} - b_i X_{aa\text{-}A,i}) - d_i X_{aa\text{-}A,i} - \alpha(i-1) X_{aa\text{-}A,i} + \alpha i X_{aa\text{-}A,i+1} \\
\dot{X}_{aa\text{-}a,1} &= \beta x_{aa} y_a - b_1 q_3 X_{aa\text{-}a,1} - d_1 X_{aa\text{-}a,1} + \alpha X_{aa\text{-}a,2} \\
\dot{X}_{aa\text{-}a,i} &= q_3(b_{i-1} X_{aa\text{-}a,i-1} - b_i X_{aa\text{-}a,i}) - d_i X_{aa\text{-}a,i} - \alpha(i-1) X_{aa\text{-}a,i} + \alpha i X_{aa\text{-}a,i+1}
\end{aligned}
\tag{59}
$$

For virgin queens we have:

$$
\begin{aligned}
\dot{x}_{AA} &= (1-q_1)\sum_i b_i(X_{AA\text{-}A,i} + \frac{1}{2} X_{Aa\text{-}A,i}) - x_{AA}(g+\beta y) \\
\dot{x}_{Aa} &= (1-q_2)\sum_i b_i(X_{AA\text{-}a,i} + \frac{1}{2} X_{Aa\text{-}A,i} + \frac{1}{2} X_{Aa\text{-}a,i} + X_{aa\text{-}A,i}) - x_{Aa}(g+\beta y) \\
\dot{x}_{aa} &= (1-q_3)\sum_i b_i(\frac{1}{2} X_{Aa\text{-}a,i} + X_{aa\text{-}a,i}) - x_{aa}(g+\beta y)
\end{aligned}
\tag{60}
$$

For males we have

$$
\begin{aligned}
\dot{y}_A &= \sum_i b_i(X_{AA\text{-}A,i} + X_{AA\text{-}a,i} + \frac{1}{2} X_{Aa\text{-}A,i} + \frac{1}{2} X_{Aa\text{-}a,i}) - h y_A \\
\dot{y}_a &= \sum_i b_i(\frac{1}{2} X_{Aa\text{-}A,i} + \frac{1}{2} X_{Aa\text{-}a,i} + X_{aa\text{-}A,i} + X_{aa\text{-}a,i}) - h y_a
\end{aligned}
\tag{61}
$$

Here g and h denote respectively the death rates of virgin queens and males. Note that virgin queens become fertilized queens after mating. Hence, the term βy is added to their death rate, where $y = y_A + y_a$ is the total abundance of males. We assume that males do not die because of the mating, but they have a very short life span. Hence it is unlikely that a male mates twice. Finally, we assume that males and females are produced at the same rate, an assumption which could easily be modified by introducing an additional parameter.

For computer simulations we add density limitation by multiplying each birth term with a factor $\phi = 1/(1+\eta X)$ where η is a constant and X is the total population size.

Figure 5 shows the equilibrium structure of the sexual model assuming that the eusocial allele, a, is recessive. This means that aa females stay with the nest with probability q, while AA and

Aa females always leave the nest. Therefore, 'full-sized' colonies are formed by aa-a queens and 'half-sized' colonies are formed by Aa-a queens. All daughters of aa-a queens are aa and stay with the nest with probability q. Half of the daughters of Aa-a queens are aa, who stay with the nest with probability q, while the other half are Aa who leave the nest. All other fertilized females are solitary, because all of their offspring leave the nest. We assume that all males leave the nest.

We use the same parameter values as for the asexual simulation. The birth rate of a solitary queen is $b_0 = 0.5$ and her death rate is $d_0 = 0.1$. The benefits of eusociality emerge once the colony reaches a critical size of $m = 3$. For all colony sizes, $i \geq m$, the birth rate of the queen is $b = 4$ and her death rate is $d = 0.01$. We make the following observation. As long as $q < 0.7$ the solitary allele is selected: all females are AA and all males are A. At $q \approx 0.7$ there is a sudden reversal. For $0.7 < q < 0.88$ the eusocial allele outcompetes the solitary one: all females are aa and all males are a. But for $q > 0.88$ we find that heterozygote females, Aa, become abundant; now many colonies are now founded by Aa females that have mated with a males. Figure 5 shows the stable equilibrium that is reached from an initial state where the eusocial allele, a, is rare. This initial condition is relevant, if we want to study the origin of eusociality in a world of solitary insects.

Figure 6 uses the same parameter values as Figure 5, but this time we show the stable equilibrium that is reached when starting from a state where the solitary allele is rare. This initial condition is relevant, if we want to study the evolutionary stability of eusociality. This time we find that the solitary allele, A, wins if $0 < q < 0.26$, while the eusocial allele, a, wins if $0.26 < q < 0.88$. For $0.88 < q < 1$ there is coexistence of the two alleles.

We have the following situation. There are three critical thresholds of q. In our numerical example they are $q_{c1} \approx 0.26$, $q_{c2} \approx 0.7$, $q_{c3} \approx 0.88$. If $0 < q < q_{c1}$ the solitary allele wins. If $q_{c1} < q < q_{c2}$ then we observe bi-stability; either the solitary or the eusocial allele wins depending on initial condition. This means each homogeneous population is stable against invasion by the other allele. If $q_{c2} < q < q_{c3}$ then the eusocial allele wins. If $q_{c3} < q < 1$ then there is coexistence between the two alleles.

12 Summary

We have proposed a model for the origin of eusociality. The basic idea is that a mutation prevents some daughters from leaving the nest. We study if this mutant allele is favored by natural selection. The fundamental consideration is the following: how are the key demographic parameters of the eusocial queen (her fecundity and her death rate) affected by the presence of workers. We find that eusociality is selected if the fecundity of the queen increases dramatically while colony size is still small. This means that a eusocial queen that is supported by a few workers must already have a significantly increased rate of oviposition, must be more successful at guarding the larvae, and having them fed until they hatch. The evolution of eusociality is also favored if the queen has a reduced death rate in the presence of workers. Both effects can arise naturally given that the eusocial queen stays at the nest, thereby reducing her risk of predation and devoting her resources to laying eggs and guarding them. Interestingly, we find that a reduced death rate alone is not sufficient for the evolution of eusociality. An increased birth rate is necessary.

A key observation of our model is that it is difficult to evolve eusociality, because we need very favorable parameters. In our numerical examples we assumed that a eusocial queen with two workers has an 8-fold increased birth rate and a 10-fold reduced death rate. For the same parameter choices, for example, a 7-fold increased birth rate would not have allowed the evolution

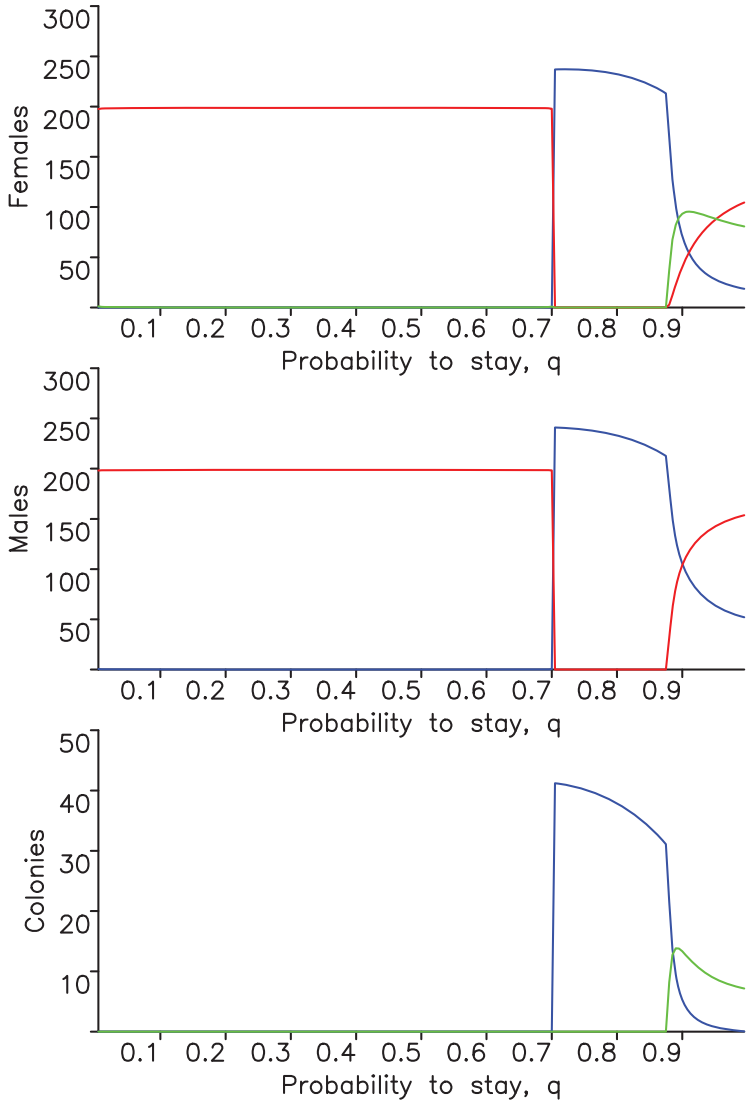

Figure 5: Emergence of eusociality in a model with haplodiploid genetics. We study the competition between a wildtype allele, A, and a recessive mutant allele, a. There are three types of females: AA (red), Aa (green) and aa (blue); aa females stay with the nest with probability q. There are two types of males: A (red) and a (blue). We simulate the system given by (59) - (61) showing the equilibrium that is reached from an initial condition where the solitary allele, A, dominates the population. For $0.7 < q < 0.88$ the eusocial allele replaces the solitary one. For $0.88 < q < 1$ there is coexistence. Parameter values: $b_0 = 0.5$, $m = 3$, $b = 4$, $\beta = 0.1$, $\eta = 0.01$; mortality rates: $d_0 = \alpha = g = h = 0.1$ and $d = 0.01$.

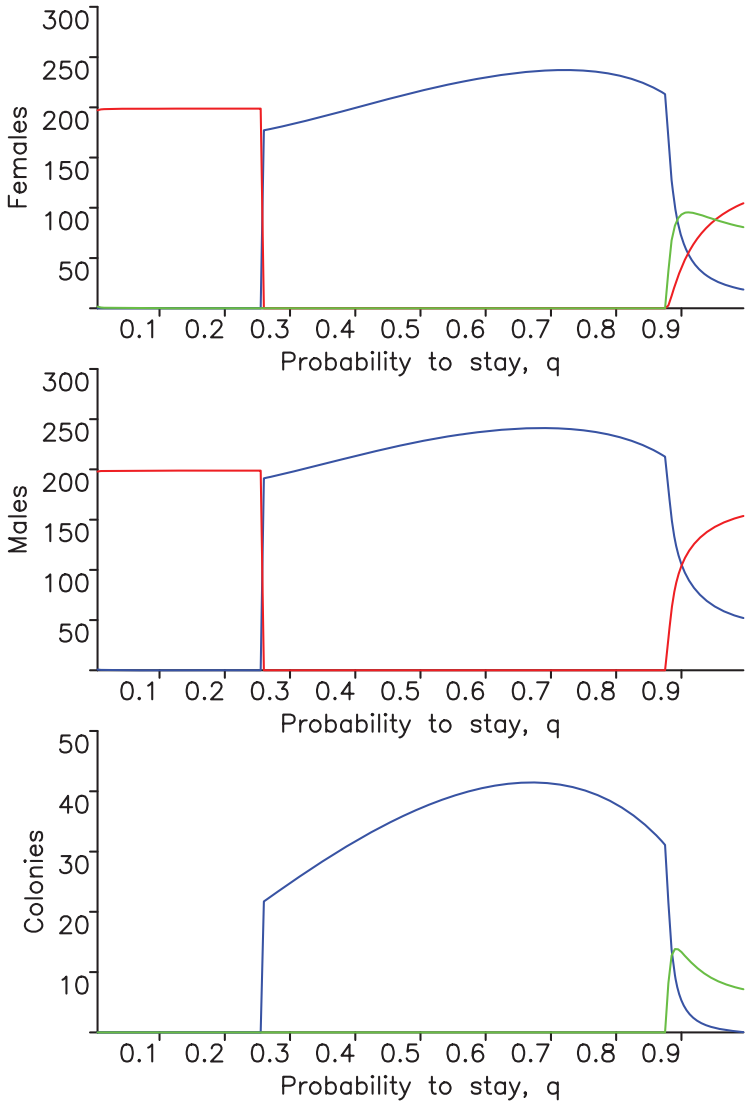

Figure 6: Evolutionary stability of eusociality in a model with haplodiploid genetics. We study the competition between a wildtype allele, A, and a recessive mutant allele, a. There are three types of females: AA (red), Aa (green) and aa (blue); aa females stay with the nest with probability q. There are two types of males: A (red) and a (blue). We simulate the system given by (59) - (61) showing the equilibrium that is reached from an initial condition where the eusocial allele, a, dominates the population. For $0.26 < q < 0.88$ the eusocial allele outcompetes the solitary one. For $0.88 < q < 1$ there is coexistence. Same parameter values as for Figure 5.

of eusociality. This observation is interesting: despite the obvious and intuitive advantages of eusociality, it is very hard for a solitary species to achieve it.

Whether or not the circumstances can be that favorable for eusociality depends on the ecological environment of the pre-eusocial species under consideration. As outlined in the main text, a key factor is whether it is possible to establish and defend a valuable nest site that is close enough to a rich food source.

Another interesting aspect of our haplodiploid model is the occurrence of bistability. It is easier to maintain eusociality than to evolve it. For a wide parameter region the eusocial allele cannot invade the solitary allele, and vice versa the solitary allele cannot invade the eusocial one (compare Figs 4 and 5). This property of the model explains in part why, even though eusociality is ecologically dominant, the condition has evolved rarely in the history of life.

Our model leads to a number of radical new suggestions for research on eusociality.

(i) Contrary to the prevailing dogma of inclusive fitness theory, it is not useful to view eusociality as an evolutionary game between workers and queens. Inclusive fitness theory claims to be a gene-centered approach, but instead it is 'worker-centered'. It puts the worker into the center of the analysis and asks the (puzzling) question: why does a worker sacrifice her reproductive potential to help raise the offspring of the queen? But if we put the gene into the center of the analysis (as we have done) then this question of altruism does not even appear. A standard selection equation determines if the eusocial allele wins over the solitary one. There is no payoff matrix, there is no evolutionary game on the decisive level of competition. By formulating a model of population genetics in the context of family structure, we find that there is no need for the inclusive fitness detour.

(ii) The queen and her workers are not engaged in a standard cooperative dilemma. The reason is that the workers are not independent agents. Their properties are determined by the alleles that are present in the queen (both in her own genome and in that of the sperm she has stored). The workers can be seen as 'robots' that are built by the queen. They are part of the queen's strategy for reproduction.

(iii) It is not useful to describe a female who stays with the colony as 'altruistic' and a female who leaves as 'selfish'. Both types of females are needed for the reproduction of the colony.

(iv) Our model does not use standard multilevel selection. There is only one level of selection, the hymenopteran colony, which is treated as an extension of the queen, whose genes are the units of selection. Selection operates between queens and solitary females or between different queens.

(v) Relatedness does not drive the evolution of eusociality. We can use our model to study the fate of eusocial alleles that arise in thousands of different presocial species with haplodiploid genetics and progressive provisioning. In some of those species eusociality might evolve, while in others it does not. Whether or not eusociality evolves depends on the demographic parameters of the queen (as discussed above), but not on relatedness. The relatedness parameters would be the same for all species under consideration.

(vi) Once eusociality has evolved, colonies consist of related individuals, because daughters stay with their mother to produce further offspring. Thus, relatedness is a consequence of eusociality, but not a cause.

(vii) Our model has clear implications for productive empirical research. The crucial measurement that needs to be performed is the effect of the size of the colony on the demographic parameters of the queen, such as her oviposition rate and average longevity. The animals most likely to yield significant data are the primitively eusocial bees and wasps, of which there are a

large number of species representing finely divided grades of evolution – including those that have reverted from eusociality to the solitary condition. There is moreover a need to identify the genetic coding of the solitary-eusocial transition, as well as to identify the phenotypic flexibility of the genes involved and the environmental pressures that drive selection.

Finally, we propose that kin selection among social insects is an apparent phenomenon which arises only when you put the worker into the center of evolutionary analysis. Kin selectionists have argued that a worker who behaves altruistically by raising the offspring of another individual, requires an explanation other than natural selection, and this other explanation is kin selection. We argue, however, that there exists a more convenient coordinate system. If the eusocial gene is in the center of the evolutionary analysis, then a standard natural selection equation determines whether or not eusociality wins. There is no paradoxical altruism that needs to be explained. The epicycles of kin selection and inclusive fitness disappear.

13 Acknowledgements

We thank Tibor Antal and Hisashi Ohtsuki for many discussions regarding the general framework presented in Part A of the online material. We thank Matthijs van Veelen for valuable insights into additive games. We thank Radu Berinde, David Haig, David Hughes, Daniel Kronauer, Erez Lieberman, David Rand and Arne Traulsen for helpful discussions. We thank Lydia Liu for help with the preparation of the manuscript.

References

Alger I and Weibull JW (2009). Kinship, incentives and evolution" (2009). forthcoming *American Economic Review*

Alexander R D (1974). The evolution of social behaviour. *Annual Review of Ecology and Systematics*, **5**, 325-383.

Anderson M (1984). The evolution of eusociality. *Annual Review of Ecology and Systematics*, **15**, 165-189.

Antal T, Ohtsuki H, Wakeley J, Taylor PD, Nowak MA (2009a). Evolution of cooperation by phenotypic similarity. *Proc Natl Acad Sci USA* **106**, 8597-8600.

Antal T, Nowak MA and Traulsen A (2009b). Strategy abundance in 2x2 games for arbitrary mutation rates. *J Theor Biol* **257**, 340-344.

Bijma P, and Wade MJ (2008). The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J Evol Biol* **21**, 1175-1188.

Boomsma JJ (2009). Lifetime monogamy and the evolution of eusociality. *Phil Trans R Soc B* **364**, 3191-3207.

Bourke AFG, Franks NR (1995). Social Evolution in Ants. Princeton NJ: Princeton University Press.

Cavalli-Sforza LL, Feldman MW (1978). Darwinian selection and "altruism". *Theor Pop Biol* **14**, 268-280.

Charnov EL (1978). Evolution of eusocial behavior: offspring choice or parental parasitism? . *J theor Biol* **75**, 451-465.

Craig R (1979). Parental manipulation, kin selection, and the evolution of altruism . *Evolution* **33**, 319-334.

Craig R (1979). Subfertility and the evolution of eusociality by kin selection . *J theor Biol* **100**, 379-397.

Crozier RH, Pamilo P (1996). Evolution of Social Insect Colonies: Sex Allocation and Kin Selection. Oxford and New York: Oxford University Press.

Doebeli M and Hauert C (2006). Limits of Hamilton's rule. *J Evol Biol* **19**, 1386-1388.

Eigen, M and Schuster, P (1977). The hyper cycle. A principle of natural self-organization. Part A: Emergence of the hyper cycle. *Naturwissenschaften* **64**, 541-565.

Fletcher JA, Zwick M, Doebeli M, Wilson DS (2006). What's wrong with inclusive fitness? *TRENDS Ecol Evol* **21**, 597-598.

Fletcher JA and Doebeli M (2009). A simple and general explanation for the evolution of altruism. *Proc R Soc B* **276**, 13-19.

Foster, K.R. et al. (2006) Kin selection is the key to altruism. *Trends Ecol Evol* **21**, 5760.

Frank SA (1998). Foundations of social evolution. Princeton University Press, Princeton, NJ.

Gadagkar R (1990). Origin and evolution of eusociality: a perspective from studying primitively eusocial wasps. *J. Genet.* **69**, 113-125.

Gadagkar R (2001). The social biology of Ropalidia marginata: Toward understanding the evolution of eusociality. Cambridge, MA: Harvard University Press.

Gardner A, West SA and Barton NH (2007). The relation between multilocus population genetics and social evolution theory. *Am Nat* **169**, 207-226.

Gardner A (2009). Adaptation as organism design. *Biol Lett* **5**, 861-869.

Gokhale C and Traulsen A (2010). Evolutionary games in the multiverse. *Proc Natl Acad Sci USA*, in press.

Goodnight, C. et al. (2008). Evolution in spatial predator-prey models and the "prudent predator: The inadequacy of steady-state organism fitness and the concept of individual and group selection. *Complexity* **13**, 2344.

Grafen, A (1979). The hawk-dove game played between relatives. *Anim Beh* **27**, 905-907.

Grafen, A (1984). Natural selection, kin selection and group selection. Chapter 3 of *Behavioural Ecology, 2nd edition (ed.s J.R. Krebs and N.B. Davies)*, **62-84**. Blackwell Scientific Publications, Oxford.

Grafen A (2007a). An inclusive fitness analysis of altruism on a cyclical network. *J Evol Biol* **20**, 2278-2283.

Grafen A (2007b). Detecting kin selection at work using inclusive fitness. *Proc R Soc B* **274**, 713-719.

Griffin, AS & West, SA (2003). Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. Science 302, 6334636.

Hamilton WD (1964). The genetical evolution of social behaviour I and II. *J Theor Biol* **7**, 1-16.

Hatchwell, BJ & Komdeur, J (2000). Ecological constraints, life history traits and the evolution of cooperative breeding. Anim Behav 59, 10791086.

Hauert C, Michor F, Nowak MA, Doebeli M (2006). Synergy and discounting of cooperation in social dilemmas. *J Theor Biol* **239**, 195-202.

Hölldobler, B. & Wilson, EO (2009). The Superorganism. W. W. Norton.

Hughes, WOH, Oldroyd, BP, Beekman, M & Ratnieks, F. L (2008). W. Ancestral monogamy shows kin selection is key to the evolution of eusociality. Science 320, 12131216.

Hunt JH (2007). The evolution of social wasps. Oxford University Press.

Karlin S and Matessi C (1983). Kin selection and altruism. *Proc R Soc B* **219**, 327-353.

Killingback T, Bieri J and Flatt T (2006). Evolution in group-structured populations can resolve the tragedy of the commons. *Proc R Soc B* **273**, 1477-1481.

Lehmann L and Keller L (2006). The evolution of cooperation and altruism – a general framework and a classification of models. *J Evol Biol* **19**, 1365-76.

Lehmann L, Keller L and Sumpter DJT (2007a). The evolution of helping and harming on graphs: the return of the inclusive fitness effect. *J Evol Biol* **20**, 2284-2295.

Lehmann L, Keller L, West S, Roze D (2007b). Group selection and kin selection: Two concepts but one process. *Proc Natl Acad Sci* **104**, 6736-6739.

Lehmann L, Ravigne V, Keller L (2008). Population viscosity can promote the evolution of altruistic sterile helpers and eusociality. *Proc R Soc B* **275**, 1887-1895.

Lieberman E, C Hauert, MA Nowak (2005). Evolutionary dynamics on graphs. *Nature* **433**, 312-316.

Linksvayer TA, Wade MJ (2005). The evolutionary origin and elaboration of sociality in the aculeate Hymenoptera: maternal effects, sib-social effects, and heterochrony. *Q Rev Biol*, **80**, 317-336.

Macfarlane A (1978). The Origins of English Individualism. Oxford: Basil Blackwell.

Macfarlane A (1992). On Individualism, *Proc Brit Acad*, **82:171-199**.

Mehdiabadi NJ, Reeve HK and Mueller UG (2003). Queens versus workers: sex-ratio conflict in eusocial Hymenoptera. *Trends Ecol Evol* **18**, 88-93.

Michener CD and Brothers DJ (1974). Where workers of eusocial hymenoptera initially altruistic or suppressed? *P Natl Acad Sci USA* **71**, 671-674.

Michod RE and Hamilton WD (1980). Coefficients of relatedness in sociobiology. *Nature* **288**, 694-697.

Moran PAP (1962). The statistical processes of evolutionary theory. Clarendon press, Oxford.

Nathanson CG, Tarnita CE, Nowak MA (2009). Calculating evolutionary dynamics in structured populations. *PLoS Comput Biol* **5**, e1000615.

Nowak M, K Sigmund (1990). The evolution of stochastic strategies in the prisoner's dilemma. *Acta Appl Math* **20**, 247-265.

Nowak MA, A Sasaki, C Taylor, D Fudenberg (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646-650.

Nowak MA (2006). Five rules for the evolution of cooperation. *Science* **314**, 1560-1563.

Nowak MA, CE Tarnita, T Antal (2010). Evolutionary dynamics in structured populations. *Phil Trans R Soc B* **365**, 19-30.

Ohtsuki H, Hauert C, Lieberman E, Nowak MA (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature* **441**, 502-505.

Ohtsuki H, Nowak MA (2006). Evolutionary games on cycles. *Proc R Soc B* **273**, 2249-2256.

Price GR (1970). Selection and covariance. *Nature* **227**, 520-521.

Price GR (1972). Extension of covariance selection mathematics. *Annals of Human Genetics* **35**, 485-490.

Queller DC (1985). Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**, 366-367.

Queller DC and Strassmann JE (1989). Measuring inclusive fitness in social wasps. *The genetics of social evolution* (p. 103-122), edited by M. D. Breed and R. E. Page, Jr. Boulder, Westview Press.

Queller DC (1994). Genetic relatedness in viscous populations. *Evol Ecol* **8**, 70-73.

Queller DC, Strassmann JE, 1998. Kin selection and social insects. *Bioscience* **48**, 165-175.

Ratnieks, FLW, Foster, KR & Wenseleers, T (2006). Conflict resolution in insect societies. Ann Rev Entomol 51, 581608.

Rousset F (2004). Genetic structure and selection in subdivided populations. Princeton, NJ: Princeton University Press.

Rousset F and Billiard S (2000). A theoretical basis for measures of kin selection in subdivided populations: finite populations and localized dispersal. *J Evol Biol* **13**, 814-826.

Roze D and Rousset F (2004). The robustness of Hamilton's rule with inbreeding and dominance: kin selection and fixation probabilities under partial sib mating. *Am Nat* **164**, 214-231.

Schneider, JM & Bilde, T (2008). Benefit of cooperation with genetic kin in a subsocial spider. Proc Natl Acad Sci USA 105, 1084310846.

Tarnita CE, T Antal, H Ohtsuki, MA Nowak (2009a). Evolutionary dynamics in set structured populations. *Proc Natl Acad Sci USA* **106**, 8601-8604.

Tarnita CE, H Ohtsuki,T Antal, F Fu, MA Nowak (2009b). Strategy selection in structured populations. *J Theor Biol* **259**, 570-581.

Taylor, PD (1989). Evolutionary stability in one-parameter models under weak selection. *Theor Popul Biol* **36**, 125143.

Taylor PD (1992). Altruism in viscous populations – an inclusive fitness model. *Evol Ecol* **6**, 352-356.

Taylor PD and Frank S (1996). How to make a kin selection argument. *J Theor Biol* **180**, 27-37.

Taylor PD, Day T and Wild G (2007a). Evolution of cooperation in a finite homogeneous graph. *Nature* **447**, 469-472.

Taylor PD, Day T and Wild G (2007b). From inclusive fitness to fixation probability in homogeneous structured populations. *J Theor Biol* **249**, 101-110.

Taylor PD and Grafen A (2010). Relatedness with different interaction configurations. *J Theor Biol* **262**, 391-397.

Taylor PD, Wild G, and Gardner A (2006). Direct fitness or inclusive fitness: How shall we model kin selection? *J Evol Biol* **20**, 296-304.

Traulsen A, Nowak MA (2006). Evolution of cooperation by multilevel selection. *Proc Natl Acad Sci USA* **103**, 10952-10955.

Traulsen A, N Shoresh, MA Nowak (2008). Analytical results for individual and group selection of any intensity. *B Math Biol* **70**, 1410-1424.

Traulsen A (2010). Mathematics of kin- and group-selection: Formally equivalent? *Evolution* **64**, 316-323.

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.

Wade MJ (1978). Kin selection: a classical approach and a general solution. *Proc Natl Acad Sci USA* **75**, 6154-6158.

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 and Zee P (2010). Multilevel and kin selection in a connected world. *Nature* **463**, E8-E9.

West SA, Griffin AS and Gardner A (2007). Evolutionary explanations for cooperation. *Curr Biol* **17**, R661-R672.

West SA, Griffin AS and Gardner A (2008) Social semantics: how useful has group selection been? *J Evol Biol* **21**, 374-385.

West SA and Gardner A (2010). Altruism, spite and greenbeards. *Science* **327**, 1341-1344.

West-Eberhard MJ (1975). The evolution of social behavior by kin selection. *Q. Rev. Biol.* **50**, 1-33.

Wild G and Traulsen A (2007). The different limits of weak selection and the evolutionary dynamics of finite populations. *J Theor Biol* **247**, 382-390.

Wild G, Gardner A and West S (2009). Adaptation and the evolution of parasite virulence in a connected world. *Nature* **459**, 983-986.

Wilson, DS (1975). A theory of group selection. *Proc Natl Acad Sci* **72**, 143-146.

Wilson DS, and Wilson EO (2007). Rethinking the theoretical foundations of socio-biology. *Q Rev Biol* **82**, 327-348.

Wilson, DS (2008). Social semantics: toward a genuine pluralism in the study of social behavior. *J Evol Biol* **21**, 368-373.

Wolf JB and Wade MJ (2001). On the assignment of fitness to parents and offspring: whose fitness is it and when does it matter? *J Evol Biol*, **14**, 347-356.

Wynne-Edwards VC (1962). Animal Dispersion in Relation to Social Behavior. Oliver and Boyd, Edinburgh.