



TOWARDS A GENERAL THEORY OF GROUP SELECTION

Burton Simon,^{1,2} Jeffrey A. Fletcher,³ and Michael Doebeli⁴

¹*Department of Mathematical and Statistical Sciences, University of Colorado at Denver 1250 Fourteenth Street, Denver, CO 80202*

²*E-mail: Burt.Simon@ucdenver.edu*

³*Systems Science Graduate Program, Portland State University Portland, OR 97207-0751*

⁴*Departments of Zoology and Mathematics, University of British Columbia, 6270 University Boulevard, Vancouver B.C.V6T 1Z4, Canada*

Received August 16, 2012

Accepted October 4, 2012

The longstanding debate about the importance of group (multilevel) selection suffers from a lack of formal models that describe explicit selection events at multiple levels. Here, we describe a general class of models for two-level evolutionary processes which include birth and death events at both levels. The models incorporate the state-dependent rates at which these events occur. The models come in two closely related forms: (1) a continuous-time Markov chain, and (2) a partial differential equation (PDE) derived from (1) by taking a limit. We argue that the mathematical structure of this PDE is the same for all models of two-level population processes, regardless of the kinds of events featured in the model. The mathematical structure of the PDE allows for a simple and unambiguous way to distinguish between individual- and group-level events in any two-level population model. This distinction, in turn, suggests a new and intuitively appealing way to define group selection in terms of the effects of group-level events. We illustrate our theory of group selection by applying it to models of the evolution of cooperation and the evolution of simple multicellular organisms, and then demonstrate that this kind of group selection is not mathematically equivalent to individual-level (kin) selection.

KEY WORDS: Evolutionary dynamics, evolution of cooperation, evolution of multicellular organisms, multilevel selection.

The concept of group selection has been around since Darwin (1871) and its importance for understanding the evolution of helping traits has been debated from the early 1960s (e.g., Wynne-Edwards 1962; Maynard Smith 1964) to the present (e.g., Wilson and Wilson 2007; West et al. 2008). The usual idea behind group selection is that genes encoding traits that are detrimental to the individual carrying those genes might still thrive on an evolutionary time scale if the trait confers an advantage to the group in which the individual is living.

One of the apparent problems with this reasoning is that the required conditions may only rarely be satisfied in natural populations, because within groups individuals carrying genes encod-

ing for the detrimental individual trait would be at a disadvantage compared to noncarriers. This powerful intuitive argument against the likelihood of group selection motivated theoreticians to construct mathematical models to test it. Thus, both John Maynard Smith (1964) and William Hamilton (1975) constructed simple models which implied that group selection was unlikely to be an important factor in natural selection, and especially in the evolution of altruism. Maynard Smith (1964) illustrated his model with reference to the colonization of haystacks on a farm by mice and models of this type are now known as “haystack” models. Key features typically include that all groups form at the same time from a randomly mixed (or mated) global population and

contain two types (e.g., cooperators and defectors), the number of groups is constant or infinite, groups start out all the same size, groups vary only in the reproduction rate of the individuals, and all groups cease to exist at the same time. Groups then contribute their individuals to the random global mixing (or mating) phase from which new groups are again formed. Models of this type are also known as Type I group selection models, in contrast to Type II models that contain more explicit group-level events, and that make the explicit assumption that groups of individuals undergoing such group-level events can be identified in a population (Mayo and Gilinsky 1987).

Interactions among individuals within groups are often modeled as public goods or prisoner's dilemma games, in which cooperation is costly to the individual, but yields nonzero-sum benefits to the group. The average individual reproductive rate is therefore higher in groups with more cooperators. However, in every group, the proportion of cooperators decreases because defectors gain a within-group selective advantage by not engaging in the costly behavior. Even though the proportion of cooperators decreases in every group, the overall fraction of cooperators in the global population can increase if the average growth rate in cooperator-dominated groups is high enough. This counterintuitive outcome is an example of "Simpson's paradox" (Simpson 1951). What the classical models of group selection really show is that Simpson's paradox is rarely strong enough to actually lead to the evolution of traits that are detrimental to individuals. However, there are other ways that group selection can work that are not properly accounted for in those models; in particular, the effects of group-level events.

Although in nature there are many types of group-level events, for example, extinction (when all the individuals in a group die), fission (when a group breaks into pieces, each piece becoming a new group), and fusion (when two groups merge into one), haystack or Type I models use a very limited and oversimplified caricature of these events. In fact, neither Maynard Smith or Hamilton recognized the importance of even the minimalist form of group-level events present in their models, that is, the frequency of global group reformation. For mathematical convenience, Maynard Smith (1964) chose to let groups exist for many generations until all initially mixed groups (heterozygous female founders in this case) fixed for defection before group reformation, and Hamilton (1975) chose the opposite extreme of having groups reform every generation. It turns out that both of these extremes blunt Simpson's paradox effect, whereas intermediate levels of group reformation frequency can more fully utilize its potential and allow cooperation to increase in the population over a greater range of parameters (Wilson 1987; Fletcher and Zwick 2004). Yet even here, it is not clear what bearing these results have on nature given that they are based on an implicit and limited notion of group-level events.

A more reasonable view of group selection emerged that analyzes the change in a population consisting of groups of individuals over a short time frame (e.g., one generation) and determines what part of the change is due to individual-level selection and what part is due to group selection. If a significant fraction of the change is due to group selection, then obviously group selection must be an important factor, at least during the analyzed time frame. This view sees group selection as a process that can be identified when it occurs by analyzing short-term population change. The analysis is typically based on the Price equation (Price 1970, 1972; Okasha 2006) or contextual analysis (Goodnight and Stevens 1997; Goodnight 2005; Okasha 2006). However, it is important to understand that the Price equation and contextual analysis are not models of population change, but methods of analyzing and interpreting population change. The basic data those techniques need must either come from an experiment, where the population change is measured, or from a mathematical model, where the population change is calculated. Our approach to group selection here is quite different. Using dynamical models of two-level population processes, we define group selection in terms of the evolutionary outcome of the process. If the outcome is due to group-level events (i.e., the outcome would be different without the group-level events), then group selection played an important role. Our view is therefore complementary to the views based on the Price equation or contextual analysis. In fact, the raw data needed for the Price equation or contextual analysis can be obtained from solutions of our models.

The purpose of this article is not to argue that certain real-world phenomena, such as human cooperation, insect eusociality, or reduced virulence in parasites, were or were not caused (or assisted) by group selection. Instead, our purpose is to present a general mathematical framework for studying two-level evolutionary dynamics, and to shed light on the mathematical distinction between individual- and group-level events. These distinctions will help to clarify exactly how group selection works over large time scales, and leads to a new definition of group selection that is valid on those time scales. In summary, our view is that group selection is about the effects of group-level events on a two-level evolutionary process. Of course, this idea is not new, and was perhaps even anticipated by Williams (1966). However, Williams and others with the same intuitive notions about group selection did not have the proper mathematical models and concepts of group-level events to work with. Our contributions here include dynamical equations of two-level population processes that are derived from first principles (i.e., the state-dependent rates at which the various model events occur) as well as a mathematically rigorous definition of group-level events and group selection based on those equations.

The last point on the nature of our models is very important, and deserves a bit of elaboration. There are no ad hoc assumptions added to our models to enable closed form solutions, or for any other reasons. The user specifies the basic model data (all the state-dependent rate functions) as accurately as possible (or desired), without restrictions, and the resulting dynamical equations can then be solved (at least in principle) to predict the future states and equilibrium configuration of the two-level population process. In the case of our partial differential equation (PDE), the equation predicts two-level population dynamics in the same sense that the “heat equation” predicts the dynamics of heat transfer in a solid. Of course our PDE cannot be applied to every possible case of interest and will not predict population change perfectly except under controlled conditions. But this is to be expected; after all, the heat equation cannot be applied to flowing liquids, radioactive solids, structures made of a small number of particles, or cases where the composition of the solid is too complicated to be specified exactly, for example, an irregularly shaped block of granite. Nevertheless, the heat equation is a physical law in the sense that if its basic assumptions are satisfied (an “infinite” number of particles arranged in a solid) and the functional form of the shape and composition of the solid is correctly specified, then it will correctly predict the dynamics of heat transfer. The mathematical form of the heat equation is always the same. It does not depend on the fine details of the example it is applied to. Likewise, our equation for two-level population dynamics has the same mathematical form regardless of the fine details of the example it is applied to, and so a rigorous and general theory of group selection can be based on it.

The rest of the article is organized as follows. In the next section, we describe a general continuous-time Markovian model of two-level population dynamics, and then derive the deterministic (PDE) model from the Markovian model. The structure of the resulting PDE has a piece corresponding to individual-level effects and a piece corresponding to group-level effects, reminiscent of the Price equation. This structure is then used in the following sections to formulate a mathematical characterization of individual- and group-level events in a model, which in turn leads to a simple and general definition of group selection in terms of evolutionary outcomes. Next, we use this new theory to study two examples of two-level population dynamics: the evolution of cooperation and the evolution of simple multicellular organisms. Finally, we summarize our theory and use it to show that two-level population processes cannot be properly analyzed in terms of individual level (e.g., inclusive) fitness measures alone, and therefore group selection, defined in terms of long-term evolutionary outcomes, is not mathematically equivalent to kin selection, as is often claimed based on short-term (one generation) dynamics.

Two Related Models of Two-Level Population Dynamics

Consider a model of a population consisting of distinct “groups” of “individuals.” The individuals may be organisms, in which case the groups are disjoint collections of organisms, for example, colonies, tribes, herds, demes. Or the groups may be simple multicellular organisms, in which case the individuals are the various cells of an organism. We assume throughout that there is a discrete set of different “traits” or “types” that characterize individuals, for example, genetic traits in organisms. Within groups, individuals are assumed to have finite lifetimes that may include (asexual) reproduction one or more times. Groups therefore change in size and composition over time due to individual-level events such as births, deaths, and migration. Although these individual-level events affect the sizes and compositions of the groups in the population, they do not directly affect the number of groups. Group-level population dynamics are governed by other kinds of events, such as fissioning (when a group breaks into pieces, each piece becoming a new group), fusion (when two groups combine to form one), or group extinction.

Specifically, we consider a population of groups of individuals, where individuals are of types $\{1, 2, \dots, k\}$. The state of a group is specified by a vector of population levels, $\vec{x} = (x_1, \dots, x_k)$, where x_i is the number of type i individuals in the group. A group in state \vec{x} is called an \vec{x} -group.

The state of the whole population at time t is

$$\theta_t(\vec{x}) = \text{number of } \vec{x}\text{-groups in the population at time } t.$$

Quantities of interest associated with the model can be calculated in terms of the state of the population. For example, the number of groups at time t and the number of type i individuals at time t are (respectively)

$$G(t) = \sum_{\vec{x}} \theta_t(\vec{x}), \quad \text{and} \quad N_i(t) = \sum_{\vec{x}} x_i \theta_t(\vec{x}). \quad (1)$$

To capture the evolutionary process, it is necessary to understand how the state of the population, $\theta_t(\vec{x})$, changes in time. The most convenient dynamical equations for our purposes here are deterministic, but only very large populations approach deterministic behavior. Our tactic is therefore to start with a stochastic (Markovian) model and derive deterministic equations by taking a limit as the populations increase to infinity.

THE MARKOVIAN MODEL

Our basic model of $\theta_t(\vec{x})$, $t \geq 0$ is a continuous-time Markov chain (Karlin and Taylor 1975). The mathematical details of this model are described in the Supporting information, where we give

a summary description of the stochastic events that occur both at the level of individuals and the level of groups.

Within groups, individuals have finite lifetimes that may include (asexual) reproduction one or more times. Type i individuals, residing in an \vec{x} -group, have birth rates $b_{ti}(\vec{x})$ and death rates $d_{ti}(\vec{x})$ at time t . Birth and death rates typically depend on interactions taking place within groups, for example, according to an evolutionary game such as a public goods game, but these rates may also depend on the state of the whole population. When a type i individual gives birth, the offspring is a type j individual with probability μ_{ij} . If the offspring is not the same type as its parent, a mutation has occurred. All offspring (initially) reside in the same group as the parent, but individuals in one group may migrate to another group. The rates that individuals migrate depend on their types, and on the states of their groups. A migrating individual chooses another group to join at random, and is accepted with a probability that depends on its type and the state of the chosen group.

At the group level, \vec{x} -groups fission at rate $f_i(\vec{x})$, and die or extinction at rate $e_i(\vec{x})$. Both these rates generally depend on the state of the group, as well as on the state of the whole population. When a group fissions, it breaks into two or more pieces that become autonomous groups. The fissioning event is statistically specified by a fissioning density, $h(\vec{u}, \vec{x})$, which is the expected number of \vec{u} -groups that result from the fissioning of an \vec{x} -group. We also assume group fusion is possible. A given \vec{x} -group and \vec{u} -group fuse at rate $\rho_i(\vec{x}, \vec{u})$. Group dispersal is also possible. A group dispersal event is a mass migration of all the individuals within the dispersing group. In the cases of fission, fusion, extinction, and dispersal, the original groups cease to exist after the event. Finally, it is also possible to include interactions between groups in our stochastic model of two-level population dynamics. For example, groups could play games against each other, with the outcome determined by the composition of the interacting groups (e.g., with group containing more cooperators being more likely to win a contest against groups containing fewer cooperators, see Supporting information).

The model of two-level population dynamics just described is similar to (but more general than) the stochastic model described in Simon (2010) and Simon et al. (2012). The mathematical details of the events in our Markovian model, and its state-dependent rate functions, are given in the Supporting information. Here it suffices to say that there is no simple way to represent the state space for $\theta_t(\vec{x})$, so the only practical way to analyze the model exactly is by (stochastic) simulation. Simulating a continuous-time Markov chain, even a complicated one like our model of two-level population dynamics, is reasonably straightforward (see Supporting information). The only practical problems occur when the populations are too large, in which case, the simulations run slowly.

THE DETERMINISTIC (PDE) MODEL

When populations of groups and/or individuals in the model are too large to efficiently simulate (and even when such simulation is feasible), the modeler instead has the option to analyze a continuous-state deterministic PDE version of the same basic model. There are a variety of reasons that a continuous-state deterministic analog of the discrete-state stochastic model may be appropriate and desirable. First of all, the PDE is derived from the discrete-state stochastic model, and so it is highly analogous to it. The PDE has an explicit mathematical form, which the simulation does not, and the PDE can be solved numerically (e.g., LeVeque 2005), yielding deterministic evolutionary trajectories and equilibrium configurations for the model. Furthermore, if the populations of groups and individuals are large enough, then (at least in special cases) the continuous-state deterministic model can be rigorously justified by laws of large numbers and other asymptotic analyses (e.g., Champagnat et al. 2006; Puhalskii and Simon 2012). In fact, we conjecture that the PDE introduced below as an approximation to the discrete-state stochastic model can be justified by a limit theorem of that type. These kinds of mathematical results are often referred to as “infinite population approximations” in the biological sciences literature. In the applied probability and queueing theory literature, they are often called “fluid limits.” There are also many analogous phenomena in physics, for example, when a large number of randomly moving particles act deterministically as a whole, as in models of gas and fluid dynamics. For models of evolutionary processes with large populations, limit theorems of these kinds show that the population dynamics in the stochastic and deterministic versions of the model will be similar. Even when populations are not very large, the continuous-state deterministic models are still useful. The solution of the PDE can serve as an approximation for the discrete population dynamics, but more importantly, the mathematical form of the PDE is very telling, and leads to an explicit characterization of group selection, as we will see.

To construct the PDE model from the simulation model, the integer-valued population variables are simply replaced by real-valued variables, and the stochastic rate functions are interpreted as deterministic rate functions. The state of a group, \vec{x} , becomes an element of \mathfrak{M}_+^k , and the state of the population, $\theta_t(\vec{x})$, becomes a density function, that is, $\theta_t(\vec{x})d\vec{x}$ is the number of groups in the population that are approximately \vec{x} -groups. Again, quantities of interest associated with the model can be calculated from $\theta_t(\vec{x})$, for example, in analogy with (1) we have

$$G(t) = \int \theta_t(\vec{x})d\vec{x} \quad \text{and} \quad N_i(t) = \int x_i \theta_t(\vec{x})d\vec{x}. \quad (2)$$

Each type of event in the model (births, deaths, fissioning, fusion, etc.) contributes a “component” mathematical piece, and the

pieces are then assembled into the PDE (eq. 7 below). The details of the construction are in the Supporting information.

THE DYNAMICAL EQUATIONS FOR THE DETERMINISTIC MODEL

In Simon (2010), it is shown that in the absence of group-level events (i.e., the only events in the model are individuals births, individuals deaths, mutation, and individual migration), the state of the population obeys the equation

$$\frac{\partial \theta_t}{\partial t}(\vec{x}) + \sum_{i=1}^k \frac{\partial(\theta_t \alpha_{ii})}{\partial x_i}(\vec{x}) = 0, \quad (3)$$

where

$$\alpha_{ii}(\vec{x}) = b_{ii}(\vec{x}) - d_{ii}(\vec{x}) \quad (4)$$

that is, the total birth rate minus the total death rate of type i individuals in an \vec{x} -group at time t . Births of type i individuals in an \vec{x} -group include type i births from type i parents, type i births from type j parents (mutations), and type i immigrants from other groups due to immigration and dispersion. Deaths of type i individuals include type i migrants that leave the group, as well as normal type i deaths. We note that (3) is a hyperbolic PDE, also called a conservation law (LeVeque 2005). In this case, the total number of groups $G(t)$ is the conserved quantity (due to the absence of group-level events). Keeping the set of groups constant of course severely restricts the potential for group selection. For example, it is well known that when the set of groups is fixed over time without new group formation, cooperative traits typically go extinct, that is, the number of cooperators in a solution of (3) converges to zero as $t \rightarrow \infty$ (in the absence of mutation). This is true for both our examples in the Examples section below.

Now consider a population of groups without individual-level births, deaths, mutation, migration, or dispersion events, so that the change in $\theta_t(\vec{x})$ in the time interval $[t, t + dt]$ is entirely due to fissioning, fusion, and extinction of groups, that is,

$$\begin{aligned} \theta_{t+dt}(\vec{x}) - \theta_t(\vec{x}) = & [\phi_t(\vec{x}) + \psi_t(\vec{x}) - (e_t(\vec{x}) \\ & + f_t(\vec{x}) + r_t(\vec{x}))\theta_t(\vec{x})]dt + o(dt), \end{aligned}$$

where $\phi_t(\vec{x})$ and $\psi_t(\vec{x})$ are the rates at which \vec{x} -groups are “born” at time t due to fissioning and fusion, and $e_t(\vec{x})$, $f_t(\vec{x})$, and $r_t(\vec{x})$ are the rates at which individual \vec{x} -groups “die” at time t due to extinction, fissioning, and fusion ($o(dt)$ is a quantity that vanishes at least to second order as dt goes to 0). Taking a limit yields,

$$\frac{\partial \theta_t}{\partial t}(\vec{x}) = g_t(\vec{x}), \quad (5)$$

where

$$g_t(\vec{x}) = \phi_t(\vec{x}) + \psi_t(\vec{x}) - (e_t(\vec{x}) + f_t(\vec{x}) + r_t(\vec{x}))\theta_t(\vec{x}). \quad (6)$$

(Details on the rate functions in (6) are provided in the Supporting information). Usually, equation (5) describes an unrealistic model of two-level population dynamics, because there are no individual-level births or deaths. However, combining (3) and (5) yields the equation for the full model of two-level population dynamics,

$$\frac{\partial \theta_t}{\partial t}(\vec{x}) + \sum_{i=1}^k \frac{\partial(\theta_t \alpha_{ii})}{\partial x_i}(\vec{x}) = g_t(\vec{x}). \quad (7)$$

It is equation (7), and not (3) or (5) that yields the most interesting and realistic two-level population dynamics.

If there is an equilibrium for (7) then $\lim_{t \rightarrow \infty} \frac{\partial}{\partial t} \theta_t(\vec{x}) = 0$, so an equilibrium density, $\theta^*(\vec{x})$, satisfies

$$\sum_{i=1}^k \frac{\partial(\theta^* \alpha_{ii}^*)}{\partial x_i}(\vec{x}) = g^*(\vec{x}), \quad (8)$$

where $\alpha_{ii}^*(\vec{x})$ and $g^*(\vec{x})$ are computed from $\theta^*(\vec{x})$, that is, in equilibrium the individual-level effects and group-level effects balance exactly.

A General Definition of Group Selection

The PDE (7) decomposes into equations (3) and (5). Equation (3) shows how the populations of group types, as described by the density function $\theta_t(\vec{x})$, change in time when groups never fission, fuse, disperse, or die of extinction. In other words, the groups retain their identities throughout time, so the only changes in the population are due to the population dynamics of individuals within the groups. Based on these observations, it is natural to interpret (3) as an equation that accounts for individual-level effects only. Likewise, equation (5) shows how $\theta_t(\vec{x})$ changes in time when there are no individual-level births, deaths, or migrations; so there are no individual-level effects. It is therefore natural to interpret (5) as an equation that accounts for group-level effects only. These insights suggest simple definitions of individual-level and group-level events and effects.

- If an event featured in a model of two-level population dynamics changes the value of one or more of $\{\alpha_{i1}(\vec{x}), \dots, \alpha_{ik}(\vec{x})\}$ (i.e., the partial derivative terms on the left-hand side of (7)), then we say the event has effects at the individual level.
- If the event changes the value of $g_t(\vec{x})$ (i.e., the right-hand side term in (7)), we say the event has effects at the group level.
- If the event changes both the α_{ii} and g_t , then the event has effects at both levels.
- If the event changes the α_{ii} , but not g_t , then it is an individual-level event.

- If the event changes g_t , but not the α_{ti} , then it is a group-level event.

These definitions carry over directly to the simulation model because there is a one-to-one correspondence between the events in the PDE model and the events in the simulation model.

Calling a feature of a model, an individual-level event does not imply that it has no effect at all at the group level. For example, as the state of a group changes due to individual-level events, the fissioning and extinction rates change, which affects the timing of those group-level events. Likewise, when a group-level event like extinction occurs, the number of individuals in the population changes. The definitions are based on the mathematical structure of the PDE, so one must be careful about reading too much into the words we chose to describe them. Perhaps it would be more accurate to say that group-level events directly affect group-level population dynamics, and only indirectly affect individual-level population dynamics; and conversely for individual-level events. In any case, the definitions are appropriate for all the kinds of events considered in this article, and would appear to be justified in general.

Any mathematical theory of group selection must be based on a particular mathematical model of two-level evolutionary population dynamics which can (at least in principle) be analyzed to determine the evolutionary dynamics of some trait. Ideally, the model allows all the important individual- and group-level features of two-level population dynamics, including at a minimum, individual-level births and deaths, and group-level fissioning (or dispersion) and extinction. In other words, there are (at least) individual- and group-level births and deaths in the model, so that explicit two-level selection can occur. In real examples, other features of two-level population dynamics may also be important, like mutation, migration, fusion, and competition and cooperation among groups, that is, group-level games. The PDE (7) that governs the population dynamics for the deterministic version of our two-level model can be used to characterize group selection, because all the aforementioned features of two-level population dynamics (and surely many others) can all be explicitly accounted for in the model, and the model can (at least in principle) be solved for explicit evolutionary trajectories and equilibrium configurations. In addition, the PDE (7) has the property that it can be decomposed into two pieces, (3) and (5), that correspond to the dynamics due to individual-level-only and group-level-only effects. By comparing (3) and (7), this decomposition can be used to determine if group-level effects are important for the unfolding evolutionary dynamics.

Definition: A trait *evolves* by group selection in a model of two-level population dynamics if it establishes itself when group-level events are present in the model, and does not establish itself in the same model when they are absent.

Roughly speaking, if the group-level events in a model are decisive in the evolution of a trait, then we say that the trait evolves by group selection. On the other hand, if the trait is likely to evolve with or without the group-level events, then one cannot say that it evolves by group selection, although group-level events may still have a profound affect on its evolution.

Definition: A trait is *assisted* by group selection in a model of two-level population dynamics if it establishes itself in the model more quickly and/or more completely when group-level events are present in the model than when they are absent.

Group-level events are removed from the PDE model (7) by setting $g_t(\vec{x}) = 0$, that is, by solving (3). In the Markovian model, the group-level events are removed by setting their rates to zero. To determine whether a trait evolves or is assisted by group selection in a model, the modeler solves (3) and (7), and compares the results. Often (as in both our examples below) the fate of a trait (e.g., a cooperative trait) in (3) is obvious (it can not evolve) so only (7) needs to be solved. In cases where group selection may be assisting the evolution of a trait, but is not decisive, both equations need to be solved to quantify the effect of the group-level events.

It is important to note that in principle, the above definitions apply to outcomes over any time horizon. In particular, our theory can also be used to assess the effects of group-level events on short-term outcomes, such as instantaneous growth rates of cooperators. Inasmuch as long-term evolutionary dynamics are viewed as the outcome of selection, whereas short-term dynamics are viewed as the process of selection, our definition of group selection thus applies to both the process and the outcome of selection.

Of course, there will always be at least some small quantitative differences between the solutions of (3) and (7), so normally we will only say that a trait is assisted by, or evolves by group selection in those cases where the differences are more dramatic, or (preferably) when the resulting two-level evolutionary dynamics are qualitatively different. As a practical matter, to determine the contribution of group selection in a given model, we start by solving (3), if necessary, and continue by solving (7) with a variety of different group-level event-rate functions, that is, by varying $g_t(\vec{x})$, to determine whether group-level events change the dynamics qualitatively.

In general, a feature in a model of two-level population dynamics can have effects at the individual and group levels. For example, the effects of dispersion show up in both $\alpha_{ti}(\vec{x})$ and $g_t(\vec{x})$, and group-level interactions (games) can also have effects at both levels. (See the Supporting information for details.) When events have effects at both levels, it is more difficult to distinguish group selection from individual-level selection, so for now, we restrict our attention to models where all the featured events have effects at either the individual level or the group level, but not both. Because individual-level births, deaths, mutation, and migration operate at the individual level only, and fissioning, fusion, and

extinction operate at the group level only, any model of two-level population dynamics featuring any combination of those events is within the scope of our present theory. In future work, it should be possible to extend the present theory of group selection to examples featuring events such as dispersion and group-level games, with effects at both levels.

Examples

In this section, we illustrate our dynamical equations and our theory of group selection with two examples. The first example is a model of the evolution of cooperation, loosely based on “traditional” multilevel models of hunter–gatherer tribes (e.g., Bowles et al. 2003). In each tribe, “defectors” have a higher birth rate than “cooperators,” but at the group level, tribes with a higher fraction of defectors are more likely to fission and/or die of extinction. The second example is a model of the evolution of a simple multicellular organism in a population containing predators. “Normal” cells reproduce faster than “sticky” cells, but sticky cells are more likely to find themselves in groups that are large enough to be safe from predators.

In both examples, there are numerous possible choices of rate functions for the individual- and group-level events, and numerous parameters associated with the rate functions that can all be independently varied. In future work, these examples can be studied in more detail, using rate functions carefully chosen to represent realistic (or otherwise interesting) cases. However, no attempt will be made to study these examples in that kind of detail here. Instead, simple generic-looking rate functions will be used, and the equations will be solved for a small number of parameter settings chosen to illustrate our main points.

THE EVOLUTION OF COOPERATION IN HUNTER–GATHERER TRIBES

Perhaps the most widely studied examples of two-level population processes are models of groups made up of “cooperators” and “defectors.” The models analyzed in Simon et al. (2012) and Simon (2010) are of this variety, and our present model is based on the same basic assumptions. The groups can be thought of as tribes of hunter–gatherers, although other examples can be represented with similar models. The basic assumptions are that:

- (i) Defectors have higher birth rates than cooperators in every tribe.
- (ii) Larger tribes, and tribes with a larger proportion of defectors, are more likely to fission.
- (iii) Smaller tribes, and tribes with a larger proportion of defectors, are more likely to die of extinction.

Fissioning events can be thought of as results of intratribe squabbling, more likely to occur in noncooperative tribes, but occurring occasionally in all large tribes. The extinction events can be thought of as results of intertribe competition, where large and cooperative tribes have an advantage over small and noncooperative tribes.

Figures 1 and 2 show examples of numerical solutions of both the stochastic and the deterministic version of our models. Figure 1a and b show the time-dependent populations of tribes, cooperators, and defectors; and Figure 2a and b show the equilibrium configurations in the stochastic and the deterministic model, respectively. Note that the two versions agree well both in regard to the evolutionary trajectory and the equilibrium configuration. This is consistent with the conjecture that the PDE model is a large population limit of (a properly scaled sequence of) simulation models. A video of a stochastic realization of the model can be found at the following link: <http://www.youtube.com/watch?v=EBW41K3xw0I&feature=youtu.be>, and a video of a PDE solution of the model can be found at the following link: <http://www.youtube.com/watch?v=87UAHkCK1qQ&feature=plcp>.

In this example, cooperation is sustained by mutation only when the group-level events (fission and extinction, in this case) are removed from the model, because defectors have higher birth rates than cooperators in every group, and equal death rates. Our definition of group selection implies that if cooperators establish themselves in the population in the long term when the group-level events are included, then the cause of their success is group selection. Our examples show that cooperators can indeed establish themselves in the equilibrium configuration due to group-level events (Figs. 1 and 2), and furthermore, a detailed analysis of $\theta_t(\vec{x})$, $t \geq 0$ shows exactly how they establish themselves, starting from the initial condition.

STICKY CELL POPULATION DYNAMICS

In the next example, we imagine a population that contains two kinds of cells: “sticky” and “normal.” The cells typically find themselves in groups consisting of cells of both types, which we will refer to as “organisms.” An organism is therefore specified by the number of sticky and normal cells it is made of, that is, an organism is of type (x, y) , where x is the number of sticky cells, and y is the number of normal cells. Over time, organisms change in size and composition due to births and deaths of sticky and normal cells. In addition, organisms occasionally fission randomly into two pieces, and occasionally two organisms will fuse into one. The population also contains “predators” that feed on organisms. The basic assumptions of the model are as follows:

- (i) Normal cells reproduce at a faster rate than sticky cells.
- (ii) Stickier organisms are less likely to fission and more likely to fuse.
- (iii) Smaller organisms are more likely to be eaten.

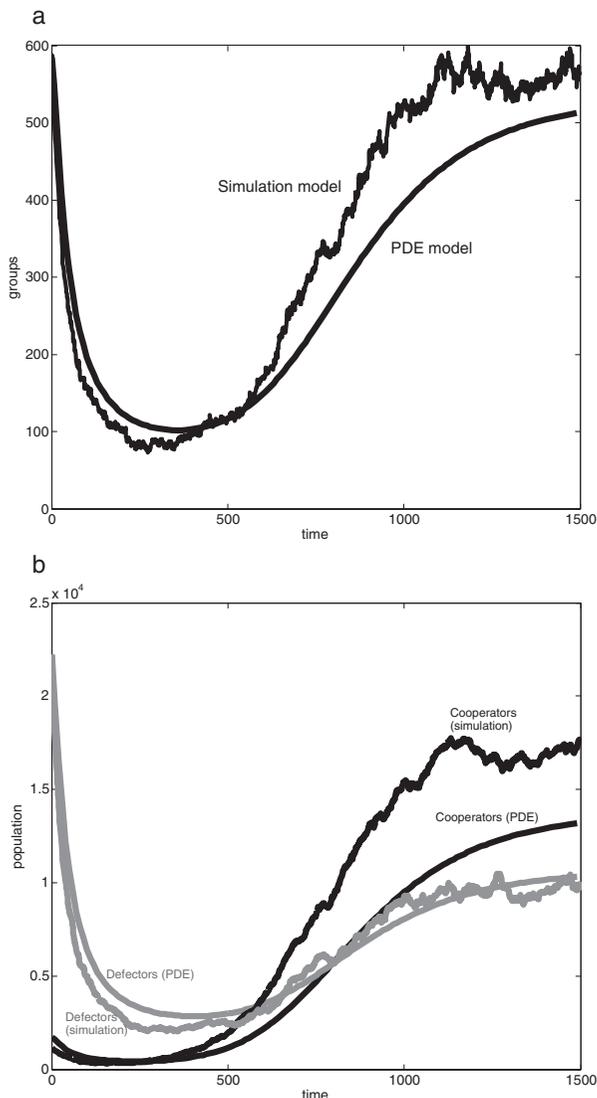


Figure 1. Population dynamics for the evolution of cooperation example, from the simulation and PDE models, showing (a) the numbers of tribes, and (b) the total numbers of cooperators and defectors. The similarity between the simulation and PDE solutions is striking. The initial state in the model consists of about 600 tribes, each made up of about 38 defectors and two cooperators. (See the Supporting information for model details.) At first, the population of tribes crashes due to fissions and extinctions of small uncooperative tribes. The remaining tribes are slightly more cooperative than the original tribes on the average, and over time the fraction of cooperators in the tribes gradually increases by group selection, allowing the population of tribes to increase until an equilibrium is reached.

As with the previous example, these assumptions are simple and reasonable, and there are many ways to construct models with these properties. The functional forms of all the rate functions used in the present example can be found in the Supporting information. The parameters of the rate functions were chosen to keep the populations at reasonable levels.

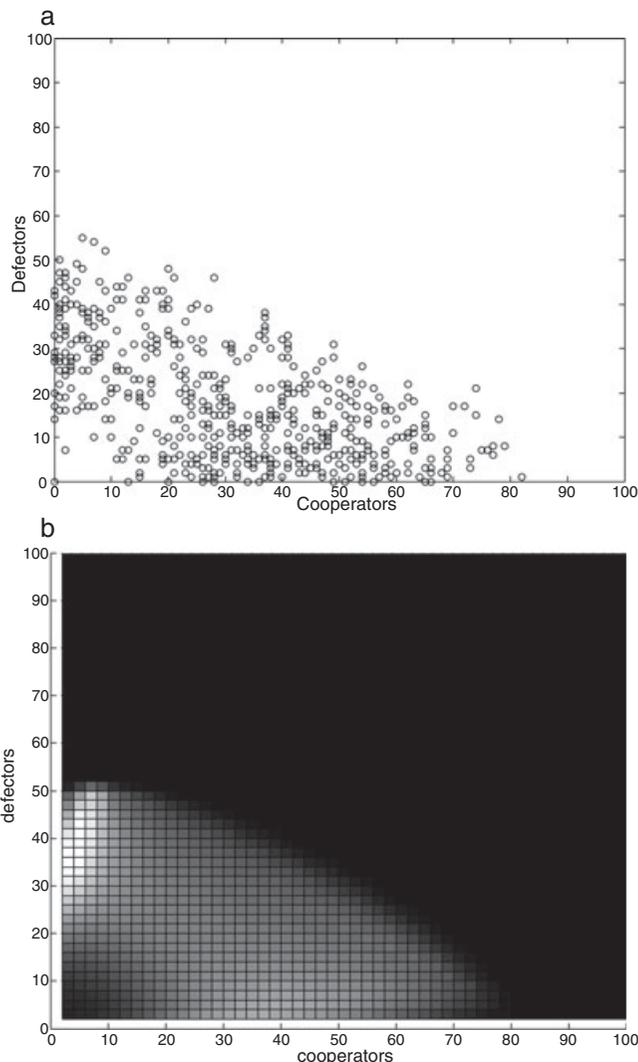


Figure 2. Equilibrium configuration for the evolution of cooperation example, from the (a) simulation and (b) PDE models. In equilibrium, the distribution of tribe types is bimodal, with a large number of tribes made up of 30–50 (mostly) defectors, and a large number of tribes made up of 40–70 (mostly) cooperators. The density is saddle shaped, so there are also tribes in transitional states. Due to the individual-level population dynamics within the tribes, cooperative groups cannot get much bigger than 80 individuals, and uncooperative tribes cannot get much bigger than 50 individuals, so there are very few tribes above the resulting diagonal line. Also, since small tribes are prone to extinction (if they do not grow quickly), there are very few of them. Again, the similarity between the simulation and PDE solutions is striking.

Because the normal cells in every organism have a higher birth rate than the sticky cells, and an equal death rate, in the absence of group-level events sticky cells gradually disappear in every organism. Our definition therefore implies that the only way stickiness can establish itself in this model is by group selection. Figures 3 and 4 show an example of the numerical

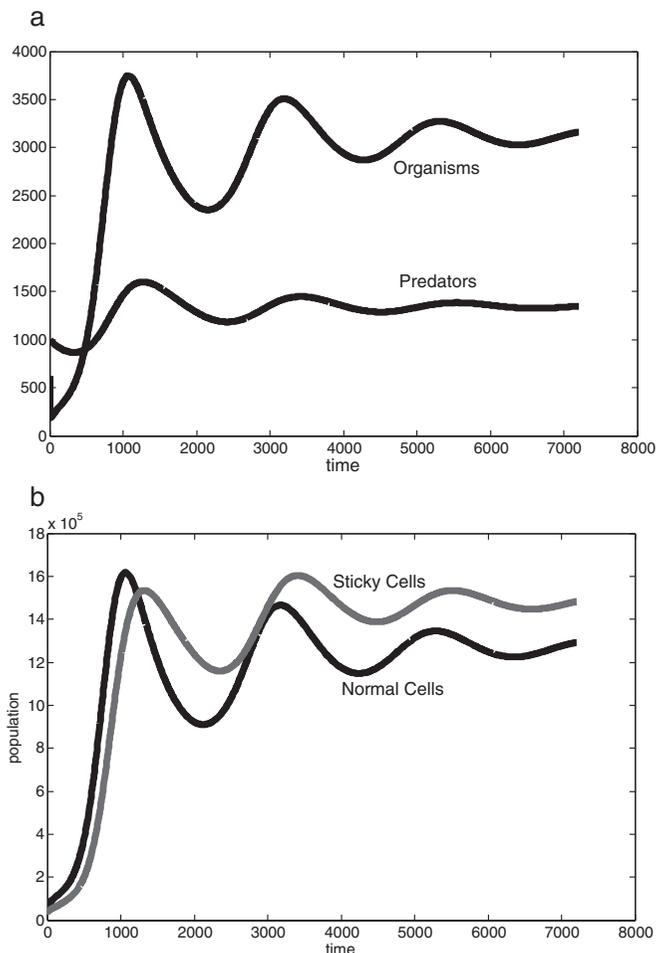


Figure 3. Population dynamics for the evolution of multicellular organisms example, from the PDE model, showing (a) the number of organisms and predators, and (b) the total number of normal and sticky cells. The initial state of the model consists of about 273 organisms, each made up of about 50 sticky cells and 100 normal cells; and about 1000 predators. (See the Supporting information for model details.) The astute reader will notice the sharp loss of organisms at $t \approx 0$. (Perhaps these original organisms suddenly found themselves in a new environment.) The sudden drop is due to organisms that fission and/or get eaten. In this environment, the original organisms that are not immediately eaten grow and/or fuse into organisms of about 600 cells. Once the initial “shock” is over, there is rapid recovery in the number of organisms, and the organisms gradually get stickier over time, by group selection. The oscillations in the populations are typical in predator/prey models.

solutions of the PDE model for the evolution of stickiness. In Figure 3, the populations of organisms, total normal cells, total sticky cells, and predators, oscillate (typical of predator/prey dynamics) until they reach equilibrium levels. Figure 4 illustrates how the state of the population of organisms changes over time. In the example shown, group selection generates

an equilibrium configuration in which stickiness is maintained (Fig. 4d).

Discussion

Essentially, the theory presented here is a mathematical formulation of Type II group selection models as envisaged by Mayo and Gilinsky (1987), who distinguished Type II models, in which birth and death events explicitly occur at both the individual level and the group level, from Type I models, in which birth and death events are only explicit at the individual level, and group properties, such as average growth rate, are only implicitly emerging from these individual level events. The basic assumption underlying Type II models is that groups of individuals undergoing group-level events, such as group fissioning and extinction, can be identified in a population. Various forms of explicit Type II models have been considered in the literature (e.g., Avilés 1986; Maynard Smith and Szathmáry 1997; Bowles et al. 2003; Traulsen and Nowak 2006); however, no general mathematical formulation of such models seems to be available. Instead, traditional descriptions of group selection are most often based on Type I models, and in particular on the Price equation (Price 1970, 1972), which, in its usual form, does not account for group-level events, and says nothing about long-term evolutionary dynamics. Nevertheless, this equation is often used to define group selection based on separating within-group and between-group effects (West et al. 2008). One example of the shortcomings of this approach is that it leads to the conclusion that group selection and kin selection (inclusive fitness) are “mathematically equivalent” (Queller 1992; Frank 1998; Lehmann and Keller 2006; Gardner 2008; but see Van Veelen et al. 2012). This has led some to conclude that group selection is not really worth studying at all (West et al. 2008). Others look at the same results and draw the opposite conclusion about the value of studying group selection (Sober and Wilson 1998; Wilson and Wilson 2007). In our opinion, these longstanding disagreements over the value of group selection are based on oversimplified models of two-level population dynamics and an inappropriate definition of group selection. (Note that the usual definition of group selection in terms of the Price equation is accepted by a wide range of researchers on both sides of the group selection debate.) Apparently, the claim that models of group selection can be recast as an analogous kin selection model (Queller 1992; Lehmann et al. 2007) is often true when only one time step is analyzed (but not always, e.g., Van Veelen et al. 2012); however, in a dynamical setting like ours, the “mathematical equivalence” between group selection and kin selection is essentially never true. This is because group selection is generally not about more-cooperative groups growing faster than less-cooperative groups, which is the characterization of group selection based on the Price equation. (Although differing group

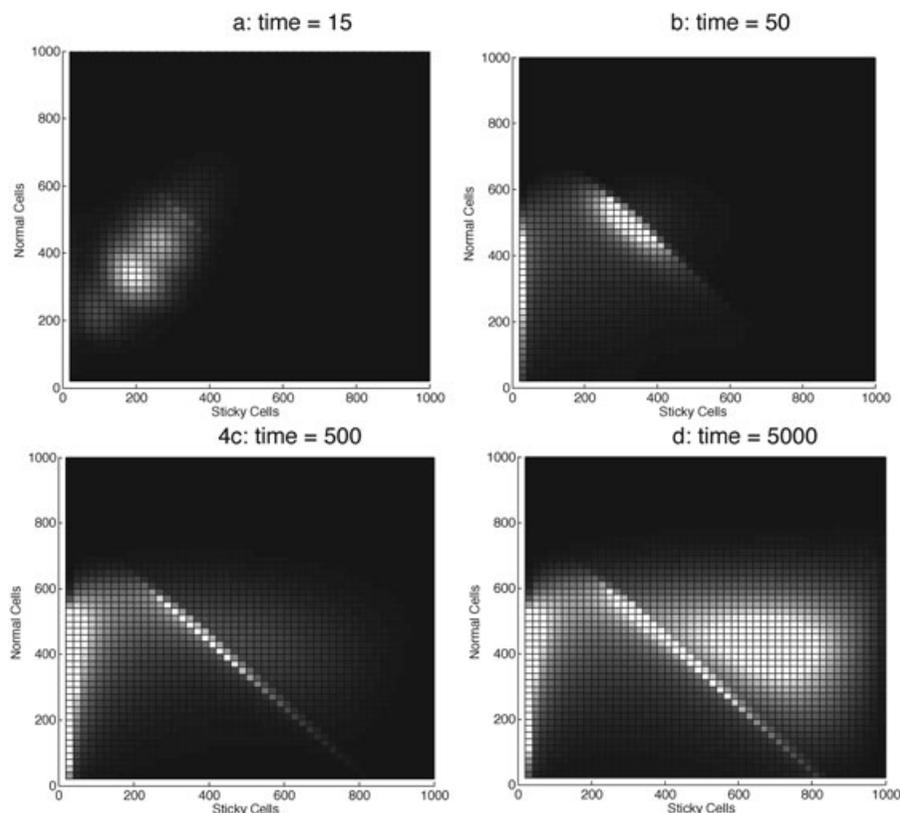


Figure 4. Dynamics of the population state density function, $\theta_t(x, y)$, at values (a) $t = 15$, (b) $t = 50$, (c) $t = 500$, and (d) $t = 5000$. In Figure 3, we see the total number of organisms, sticky cells, and normal cells, as functions of time, but we cannot tell from that information what the assortment of sticky and normal cells is like. (a) The density right after the initial population crash. We can see that the original groups that were not eaten have grown and begun to fuse (resulting in the “ripples” in the density). (b) The first hints of a stable configuration are apparent. There are a significant number of organisms with enough sticky cells to reach the equilibrium size of about 800 cells. Due to the fissioning of those organisms, and the gradual increase of the fraction of normal cells within each organism, there are also a large number of smaller organisms made up of mostly normal cells. (c) A small number of large (mostly) sticky organisms have emerged, and in (d), the equilibrium has been reached. The equilibrium consists of a complicated combination of relatively small organisms made up of mostly normal cells, organisms with various ratios of sticky and normal cells at the equilibrium size of about 800 cells, and large organisms from fusion events. The large organisms are prone to fissioning due to their size, and also gradually shrink because they are above the equilibrium size. Small- to mid-sized sticky organisms that occasionally result from the fissioning of larger groups, replenish the large sticky organisms as they gradually become less sticky, and this keeps the complicated configuration stable in time. Note that the brightness levels of the density functions are relative to the volume of the whole function. From Figure 3, one can see that the total volume of the density in (a) (for example) is much less than the total volume of the other density functions.

growth rates can play a role in group selection.) This fact is exemplified by our second example, in which sticky cells are less fit than normal cells within a group (due to their lower birth rates), and the average growth rate of individuals in stickier groups is also lower than the average growth rate of individuals in less sticky groups. The “Price fitness” of stickiness is therefore lower at both the individual and group levels, yet stickiness thrives in the model in the long term due to group-level events, that is, stickiness evolves by group selection. To reiterate, the success of sticky cells cannot be explained at the individual (within-group) level, nor in terms of stickier groups (organisms) growing faster (i.e., Simpson’s paradox), but is easily characterized in terms of group-level events: stickier groups tend to be bigger (despite their

slower growth rates) due to less fission events and more fusion events, and therefore suffer less extinction events (consumption by predators).

In spite of the arguments just given, some may still argue that at least in theory, group selection is explainable in terms of individual-level (inclusive) fitness measures, if inclusive fitness is properly defined. It is therefore important to consider the question of whether every event in a model of two-level population dynamics can be (at least in principle) considered an individual-level event, and whether this reductionist approach is useful. Our mathematical framework allows a precise answer to this question. The reductionist approach is possible if one can “absorb” $g_t(\vec{x})$ on the right-hand side of (7) into $\{\alpha_{t1}(\vec{x}), \dots, \alpha_{tk}(\vec{x})\}$ appearing

on the left-hand side, in a reasonable way, that is, if one can find functions $\beta_{ri}(\vec{x})$, $i = 1, \dots, k$ satisfying

$$\sum_{i=1}^k \frac{\partial(\theta_r \beta_{ri})}{\partial x_i}(\vec{x}) = \sum_{i=1}^k \frac{\partial(\theta_r \alpha_{ri})}{\partial x_i}(\vec{x}) - g_r(\vec{x}). \quad (9)$$

Under suitable conditions there may be solutions to (9), meaning that (7) can be rewritten as

$$\frac{\partial \theta_r}{\partial t}(\vec{x}) + \sum_{i=1}^k \frac{\partial(\theta_r \beta_{ri})}{\partial x_i}(\vec{x}) = 0, \quad (10)$$

which has the same form as (3). Equation (10) looks like (3), but has the same evolutionary trajectories as (7), and this seems to imply that everything occurs at the individual level via the functions β_{ri} . The functions β_{ri} can be thought of as a general kind of “inclusive fitness” that takes group-level events into account. If the functions α_{ri} and β_{ri} were interchangeable in some sense, then there would be some validity to the idea that the two-level and individual-level-only approaches are equivalent. However, to solve (10), equation (7) has to be solved first, because $\theta_r(\vec{x})$ appears in (9). On the other hand, (7) can obviously be solved without reference to (9), so the two-level approach is self-contained whereas the individual-level-only approach is not. Another way to say this is that the reductionist approach is not “dynamically sufficient” because a dynamical analysis via (10) requires the solution of (7) first. The approaches are therefore not mathematically equivalent. More importantly, from an informational point of view, the components $g_r(\vec{x})$ and $\alpha_{ri}(\vec{x})$ in (7) have meaningful interpretations in the model, so their values can be manipulated by the modeler in a purposeful manner, whereas the functions $\beta_{ri}(\vec{x})$, $i = 1, \dots, k$, in (10) must be obtained indirectly from (9), and they have no useful (intuitive) interpretation on their own. Equation (10) is untenable, and so the distinction between individual-level and group-level events based on (7) is real. The conclusion is that from both an aesthetic and practical point of view, models of two-level population dynamics should be formulated in terms of both individual- and group-level events, and individual-level fitness measures should only be used to model within-group population dynamics. The reductionist approach, which attempts to define and calculate individual-level (e.g., inclusive) fitness measures in such a way that they account for the entire two-level population process, obscures the real process and makes a dynamical analysis next to impossible. Such a reductionist approach would become even less feasible if more than two levels of selection were considered, that is, more than two levels at which birth and death events occur, an extension of our model that would seem to be possible in principle.

To summarize, in this article we have

- Derived dynamical equations for two-level population processes from first principles (the basic state-dependent

event-rate functions) that are “inevitable” in the sense that their basic mathematical form is the same, regardless of the fine details of the example under study.

- Defined individual- and group-level events in terms of where their effects show up in the PDE (7).
- Defined what it means for a trait to evolve by group selection or be assisted by group selection, explicitly in terms of the solutions of our dynamical equation. The definition is concerned with the evolutionary outcome of the two-level population process.
- Applied the new theory to two examples: the evolution of cooperation and the evolution of simple multicellular organisms.
- Demonstrated that the new theory is mathematically distinct from inclusive fitness theory, because the latter is not “dynamically sufficient” for two-level population processes.

ACKNOWLEDGMENTS

We thank C. Goodnight, J. Wilkins, and an anonymous referee for helpful comments on the manuscript, and A. Nielsen for his Matlab programming that produced our figures and animations.

LITERATURE CITED

- Avilés, L. 1986. Sex-ratio bias and possible group selection in the social spider *Anelosimus eximius*. *Am. Nat.* 128:1–12.
- Bowles, S., J.-K. Choi, and A. Hopfensitz. 2003. The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* 223:135–147.
- Champagnat, N., R. Ferriere, and S. Meleard. 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theor. Popul. Biol.* 69:297–321.
- Darwin, C. 1871. *The descent of man, and selection in relation to sex*. Appleton and Company, New York.
- Fletcher, J. A., and M. Zwick. 2004. Strong altruism can evolve in randomly formed groups. *J. Theor. Biol.* 228:303–313.
- Frank, S. A. 1998. *Foundations of social evolution* Princeton Univ. Press, Princeton, NJ.
- Gardner, A. 2008. The Price equation. *Curr. Biol.* 18:R198–R202.
- Goodnight, C. J. 2005. Multilevel selection: the evolution of cooperation in non kin groups. *Popul. Ecol.* 47:3–12.
- Goodnight, C. J., and L. Stevens. 1997. Experimental studies of group selection: what do they tell us about group selection in nature. *Am. Nat.* 150:S59–S79.
- Hamilton, W. D. 1964. The genetical evolution of social behavior I and II. *J. Theor. Biol.* 7:1–52.
- . 1975. *Innate social aptitudes of man: an approach from evolutionary genetics*. Pp. 133–155 in R. Fox, ed. *Biosocial anthropology*. John Wiley and Sons, New York.
- Karlin, S., and H. M. Taylor. 1975. *A first course in stochastic processes*. 2nd ed. Academic Press, London, UK.
- Lehmann, L., and L. Keller. 2006. The evolution of cooperation. A general framework and a classification of models. *J. Evol. Biol.* 19:1365–1376.
- Lehmann, L., L. Keller, S. West, and D. Roze. 2007. Group selection and kin selection: two concepts but one process. *Proc. Natl. Acad. Sci. USA* 104:6736–6739.
- LeVeque, R. J. 2005. *Numerical methods for conservation laws*. 2nd ed. Birkhauser, Basel, Switzerland.

- Marlowe, F. W. 2005. Hunter-gatherers and human evolution. *Evol. Anthropol.* 14:54–67.
- Maynard Smith, J. 1964. Group selection and kin selection. *Nature* 201:1145–1147.
- Maynard Smith, J., and E. Szathmáry. 1997. *The major transitions in evolution.* Oxford Univ. Press, New York.
- Mayo, D. G., and N. L. Gilinsky. 1987. Models of group selection. *Philos. Sci.* 54:515–538.
- Okasha, S. 2006. *Evolution and the levels of selection.* Oxford Univ. Press, Oxford, UK.
- Price, G. R. 1970. Selection and covariance. *Nature* 277:520–521.
- Price, G. R. 1972. Extension of covariance selection mathematics. *Ann. Hum. Genet.* 35:485–490.
- Puhalskii, A. and B. Simon. 2012. Discrete evolutionary birth-death processes and their large population limits. *Stoch. Models* 28:399–412.
- Queller, D. C. 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139:540–558.
- Simon, B. 2010. A dynamical model of two-level selection. *Evol. Ecol. Res.* 12:555–588.
- Simon, B., J. A. Fletcher, and M. Doebeli. 2012. Hamilton's rule in multi-level selection models. *J. Theor. Biol.* 299:55–63.
- Simpson, E. H. 1951. The interpretation of interaction in contingency tables. *J. R. Stat. Soc. B* 13:238–241.
- Sober, E., and D. S. Wilson. 1998. *Unto others.* Harvard Univ. Press, Cambridge, MA.
- Traulsen, A., and M. A. Nowak. 2006. Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci. USA* 103:10952–10955.
- van Veelen, M., J. García, M. W. Sabelis, and M. Egas. 2012. Group selection and inclusive fitness are not equivalent; the Price equation vs. models and statistics. *J. Theor. Biol.* 299:64–80.
- West, S. A., A. S. Griffin, and A. Gardner. 2008. Social semantics: how useful has group selection been? *J. Evol. Biol.* 21:374–385.
- Williams, G. C. 1966. *Adaptation and natural selection.* Princeton Univ. Press, Princeton, NJ.
- Wilson, D. S. 1987. Altruism in mendelian populations derived from sibling groups: the Haystack model revisited. *Evolution* 41:1059–1070.
- Wilson, D. S., and E. O. Wilson. 2007. Rethinking the theoretical foundation of sociobiology. *Q. Rev. Biol.* 82:327–348.
- Wynne-Edwards, V. C. 1962. *Animal dispersion in relation to social behaviour.* Oliver & Boyd, Edinburgh.

Associate Editor: C. Goodnight

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Supplementary Material. Towards a General Theory of Group Selection.

Towards a General Theory of Group Selection: Supplementary Material

Burton Simon*, Jeffrey A. Fletcher** & Michael Doebeli***

* Department of Mathematical and Statistical Sciences, University of Colorado at Denver
1250 Fourteenth Street, Denver, CO 80202

** Systems Science Graduate Program, Portland State University
Portland, OR 97207-0751

*** Departments of Zoology and Mathematics, University of British Columbia,
6270 University Boulevard, Vancouver B.C. Canada, V6T 1Z4

September 27, 2012

1 Introduction

These supplementary notes contain the mathematical details needed to construct the stochastic and deterministic models described in section 2 of the main paper, and also the specific rate functions used in the numerical examples in section 4.

2 Construction of the stochastic model

The stochastic model is specified by the state-dependent rate functions for the individual-level and group-level events featured in the model, and by a rule that shows how the state changes when one of the events occur. The following subsections describe the general forms of the rate functions for a variety of events discussed in the main text. The specific forms used in the examples are given here in sections 3 and 4.

2.1 Individual births, deaths, and mutation via evolutionary birth-death processes

The model we use for individual-level population dynamics within the groups is an evolutionary birth-death process (Simon 2010, Simon et al 2012, Puhalskii and Simon 2012). An evolutionary birth-death process is a continuous time Markov chain on Z_+^k that assigns birth and death rates to each individual based on its type, and on the state of the group it resides in. Random mutation is also allowed in the model.

Denote the birth and death rates of type i individuals in an \vec{x} -group by $b_i(\vec{x})$ and $d_i(\vec{x})$, respectively. Thus, in an \vec{x} -group, a type i individual will give birth to an offspring in a time interval of length dt with probability $b_i(\vec{x})dt + o(dt)$, and a type i individual will die in an interval of length dt with probability $d_i(\vec{x})dt + o(dt)$ (where $o(dt)$ is quantity that vanishes at least to second order as dt goes to 0). In general, $b_i(\vec{x})$ and $d_i(\vec{x})$, $i = 1, \dots, k$ may depend on how type i individuals fare in interactions within the group (e.g., a public goods game), the size of the group, $N(\vec{x}) = \sum x_i$, and/or other factors. We assume asexual reproduction, and that when an individual is born, it resides in the same group as its parent. Let μ_{ij} be the probability that a type i birth event results in a type j individual. We will call μ the mutation matrix. (If $\mu = I$, i.e., the identity matrix, then there is no mutation in the model.)

In general, the total rate that type i individuals are born in an \vec{x} -group is

$$B_i(\vec{x}) = \sum_{j=1}^k b_j(\vec{x})x_j\mu_{ji}, \quad (1)$$

and the total death rate of type i individuals in an \vec{x} -group is

$$D_i(\vec{x}) = d_i(\vec{x})x_i. \quad (2)$$

2.2 Individual migration

We will use the model of migration described in Simon (2010). Individuals move independently, and the rate at which they migrate from their current group depends on their type and the state of that group, and is denoted by $r_i(\vec{x})$. When an individual migrates from its group, it chooses at random another group in the population to try to join. The chosen group then accepts the immigrant with a probability that depends on its state, and the migrating individual's type, i.e., the probability that an \vec{x} -group accepts a migrating type i individual is $q_i(\vec{x})$. Migrants that are rejected by their chosen group are assumed to die. From a mathematical point of view, the effect of migration is to increase the individual birth rates in the groups to account for migrating individuals that join the group, and to increase the individual death rates in the groups to account for individuals migrating away. A migrating type i individual is the same as a type i death as far as the state of the group is concerned, so in a model with migration, the total death rate in an \vec{x} -group is

$$D_i^*(\vec{x}) = D_i(\vec{x}) + r_i(\vec{x})x_i. \quad (3)$$

An immigrating type i individual is the same as a type i birth as far as the state of the group is concerned, so in a model with migration, the total birth rate in an \vec{x} -group at time t is

$$B_i^*(\vec{x}) = B_i(\vec{x}) + q_i G(t)^{-1}(\vec{x}) \sum_{\vec{u}} \theta_t(\vec{u}) r_i(\vec{u}) u_i. \quad (4)$$

2.3 Extinction of groups

Of the group-level events considered here, group extinction is the simplest to model because there is only one kind of extinction event: the instantaneous death of all the members of the group. All that is needed to specify the extinction process is a rate function, $e_t(\vec{x})$, i.e., the probability an \vec{x} -group dies by extinction during the time interval $[t, t + dt]$ is $e_t(\vec{x})dt + o(dt)$. The total extinction rate of \vec{x} -groups in the population at time t is $e_t(\vec{x})\theta_t(\vec{x})$. In general, $e_t(\vec{x})$ depends on the state of the environment at time t .

2.4 Group fissioning

When a group fissions it breaks up into pieces, where the pieces consist of individuals from the original group. Thus, the combined number of type i individuals in all the pieces of a fissioned \vec{x} -group cannot exceed x_i . If the total is less than x_i then the interpretation is that some of the type i individuals died during the fissioning event. If the fissioned pieces each become a new group then the process is called *asexual group fissioning*, Okasha (2006), since each group born this way descends from a single parent group. A kind of *sexual group fissioning* will be discussed below.

In our models, fissioning is specified by two functions. The rate function, $f_t(\vec{x})$, determines the rate at which \vec{x} -groups fission at time t , and the fissioning density function, $h_{\vec{u}}(\vec{x})$, which describes the fissioning event statistically by giving the expected number of \vec{u} -groups produced by the fissioning of an \vec{x} -group. In general, $f_t(\vec{x})$ depends on the state of the environment at time t .

Perhaps the simplest kind of fissioning is a random fracture into two pieces, where the probability of ℓ type i individuals in the first piece is $1/(x_i + 1)$, $\ell = 0, 1, \dots, x_i$, and the second piece contains the remaining individuals from the group. In this case,

$$h_{\vec{u}}(\vec{x}) = 2 \prod_{i=1}^k \frac{1}{x_i + 1}, \quad (5)$$

i.e., every \vec{u} -group is equally likely (as long as $\vec{u} \leq \vec{x}$). We will call this the uniform fissioning density. Both our examples in section 5 will utilize the uniform fissioning density. Of course there are many other kinds of fissioning densities; for example, a density corresponding to individuals “flipping coins” to determine which piece to join is used in Simon (2010) and Simon et al. (2012).

The rate that \vec{u} -groups are born due to the fissioning of larger groups at time t is

$$\phi_t(\vec{u}) = \sum_{\vec{x} > \vec{u}} \theta_t(\vec{x}) f_t(\vec{x}) h_{\vec{x}}(\vec{u}). \quad (6)$$

At the group level, $\phi_t(\vec{x})$ and $f_t(\vec{x})\theta_t(\vec{x})$ (respectively) are the birth and death rates of \vec{x} -groups due to fissioning at time t .

2.5 Fusion of groups

When two groups fuse, they join together to form a single new group. The two original groups cease to exist, and the new group assumes its own identity. Some individuals from the original groups may die during the fusing process, but the fused group otherwise contains all the individuals from the original groups. For simplicity we will assume here that all the individuals survive fusing events.

Define $\rho_t(\vec{x}, \vec{y})$ be the rate that a given \vec{x} -group and \vec{y} -group, alive in the population at time t , fuse. The total rate that a given \vec{x} -group fuses with some other group at time t is therefore

$$r_t(\vec{x}) = \sum_{\vec{y}} \theta_t(\vec{y}) \rho_t(\vec{x}, \vec{y}), \quad (7)$$

which can be interpreted as the death rate of \vec{x} -groups at time t due to fusion. \vec{x} -groups are formed by the fissioning of two smaller groups at rate

$$\psi_t(\vec{x}) = \frac{1}{2} \sum_{\vec{u} \leq \vec{x}} \theta_t(\vec{u}) \theta_t(\vec{x} - \vec{u}) \rho_t(\vec{u}, \vec{x} - \vec{u}). \quad (8)$$

At the group level, $\psi_t(\vec{u})$ and $r_t(\vec{u})$ are (respectively) the birth and death rates of \vec{u} -groups due to fusion at time t .

2.6 Fission-Fusion

In some examples, a piece of a group may break off, and instead of becoming a new group in its own right, it joins (or attempts to join) another group. This process is an example of sexual group fissioning (Okasha 2006), since groups in the population may be descended from pieces of more than one parent group. If there is typically some time elapsed between the breaking apart of a group, and the subsequent fate of the pieces, then one can model this process as a fissioning event followed by a fusion event.

2.7 Group dispersion

In some examples, some or all of the individuals in a group may suddenly disperse and attempt to join other groups. This kind of event can be thought of as a mass migration, or as a special kind of fission-fusion event. However, from a mathematical point of view it appears best to think of dispersion as a mass migration, and formulate the model that way.

Suppose \vec{x} -groups undergo a dispersion event at rate $\lambda_t(\vec{x})$. The total rate that type i individuals disperse in the population at time t is therefore $\sum_{\vec{x}} \theta_t(\vec{x}) \lambda_t(\vec{x}) x_i$. As in our model of migration, our model of dispersion assumes that dispersing type i individuals choose a group in the population to join at random, and is accepted by an \vec{x} -group with probability $q_i(\vec{x})$. In a model with dispersion, the total birth rate of type i individuals in an \vec{x} -group at time t is therefore increased by

$$\hat{b}_{it}(\vec{x}) = G(t)^{-1} q_i(\vec{x}) \sum_{\vec{u}} \theta_t(\vec{u}) \lambda_t(\vec{u}) u_i \quad (9)$$

due to dispersion events in other groups.

2.8 Group-Level Games

In an evolutionary birth-death process the birth and/or death rates of the individuals in a group are often functions of the outcomes of interactions with other group members (modeled as the expected payoff they get in a game played against the other group members). For example, in models of the evolution of cooperation, the individual-level games played within the groups are typically prisoner's dilemmas or public goods games. In a model of two-level population dynamics, interactions between groups in the population may be characterized by group-level games. For example, in Simon (2010) the interactions between hunter-gatherer tribes are modeled as games of "chicken" (also known as "snowdrift"). Let $\xi(\vec{x}, \vec{y})$ be the payoff to an \vec{x} -group in a game against a \vec{y} -group. One possible measure of how \vec{x} -groups perform in the game at time t would be its average payoff against a randomly selected group, i.e.,

$$\omega_t(\vec{x}) = G(t)^{-1} \sum_{\vec{y}} \theta_t(\vec{y}) \xi(\vec{x}, \vec{y}).$$

The outcomes of these games may affect the rates of group-level events and/or individual-level events. For example, groups that fare poorly may be more likely to die of extinction, and groups that do well may have increased birth rates for their members.

2.9 Solving the stochastic model by simulation

By amalgamating all the sub-models mentioned above, one obtains a stochastic model in the form of a continuous-time Markov chain, albeit with a fairly complicated state space. In order to specify the state of the process at time t , one must know how many groups there are, and the state of each of those groups. The state of each group is specified by the vector of populations of each individual type, i.e., an element of Z_+^k . If there are ℓ groups in the population at time t , then the set of possible configurations is $(Z_+^k)^\ell$. Since ℓ can potentially be any nonnegative integer, the state space for the discrete-stochastic model is $(Z_+^k)^\infty$. Except in very special cases, when the state space of a continuous-time Markov chain is complicated like this, the only way to study it in detail is by simulation.

In order to efficiently simulate a continuous-time Markov chain, one exploits the basic mathematical properties of the exponential probability distribution which governs the times between events. At each point in time, there are certain events that may happen next, e.g., a birth or death of an individual in one of the groups, a fissioning event or extinction event in one of the groups, etc. The time until the

next event is exponentially distributed, and the rate associated with that exponential random variable is the sum of the rates associated with the various possible next events. To determine the time and identity of the next event, the simulation generates exponential random variables associated with each possible next-event, and chooses the smallest of them. Once the time and identity of the next event is determined, the state of the process is updated (e.g., using the event descriptions given in this section), and the procedure is repeated. See, for example, Ross (2010) for details on simulating a continuous-time Markov chain. It is not difficult to write a program to simulate the continuous-time Markov chain as just described. However, some events (like fusion and dispersal) seem to be inherently more computationally intensive than others (like individual births and deaths, or fissioning). The number of computations also increases with the sizes of the populations in the model, so the feasibility of a simulation analysis depends on the size and kinds of events featured in the model being studied. An example of a stochastic simulation can be found at the following link: <http://www.youtube.com/watch?v=EBW41K3xw0I&feature=youtu.be>. The simulation shown is one particular realization of the stochastic model for the evolution of cooperation presented in section 4.1. An example of a PDE simulation can be found at the following link: <http://www.youtube.com/watch?v=87UAHkCK1qQ&feature=plcp>. The simulation shown is one particular solution of the PDE model for the evolution of cooperation presented in section 4.1.

Although one does not get “formulas” from a simulation of a continuous-time Markov chain, one does get sample paths of the evolutionary trajectories, which are very informative, and statistical properties of the evolutionary trajectories can be estimated (to any desired degree of precision) by repeating the simulations. When models are too big to efficiently simulate, one can study our models of two-level population dynamics in a completely different way, using a continuous-deterministic analog of the simulation model.

3 Constructing the component pieces of the PDE

The deterministic (PDE) model is derived directly from the stochastic model. Model parameters that appear in the PDE are the same parameters that appear in the stochastic model. In order to derive the continuous model, it is only necessary to interpret the population variables, $\vec{x}(t)$, as continuous, and the stochastic rate functions e.g., (1), (2), (6), (7), (8), as deterministic rate functions. In particular, the component pieces of the PDE can be derived from the component pieces of the discrete-stochastic model, as follows.

3.1 Individual births, deaths, mutations, and migration

From (3) and (4) we obtain

$$\alpha_{ti}(\vec{x}) = B_{ii}^*(\vec{x}) - D_{ii}^*(\vec{x}), \quad i = 1, \dots, k, \quad (10)$$

where \vec{x} is interpreted as a vector of continuous variables.

3.2 Fissioning, fusion, and extinction of groups

The uniform fissioning density (5) becomes

$$h_{\vec{u}}(\vec{x}) = \frac{2}{x_1 x_2 \cdots x_k}. \quad (11)$$

The rate at which \vec{u} -groups are born due to larger groups fissioning at time t has the form

$$\phi_t(\vec{u}) = \int_{\vec{x}} \theta_t(\vec{x}) h_{\vec{x}}(\vec{u}) d\vec{x}. \quad (12)$$

For the model of fusion, the rate at which \vec{u} -groups are created from a fusing event has the form

$$\psi_t(\vec{u}) = \frac{1}{2} \int_{\vec{x}} \theta_t(\vec{x}) \theta_t(\vec{u} - \vec{x}) \rho_t(\vec{x}, \vec{u} - \vec{x}) d\vec{x}, \quad (13)$$

and the death rate of \vec{x} -groups due to fusion is

$$r_t(\vec{x}) = \int_{\vec{y}} \theta_t(\vec{y}) \rho_t(\vec{x}, \vec{y}) d\vec{y},$$

where $\rho_t(\vec{x}, \vec{y})$ is the rate a given \vec{x} -group and \vec{y} -group fuse. The rates that \vec{x} -groups die due to extinction, fissioning, and fusion at time t are therefore $e_t(\vec{x})\theta_t(\vec{x})$, $f_t(\vec{x})\theta_t(\vec{x})$, and $r_t(\vec{x})\theta_t(\vec{x})$, respectively.

3.3 Dispersion

Dispersion events have two aspects with qualitatively different effects on the state of the population. The first aspect is the dispersion of individuals into the population that independently look to join other groups. The second aspect is the sudden vanishing of the parent group.

From (9) we see that dispersion affects the total birth rates of individuals in the groups because a dispersing individual that joins a group is just like a birth. Thus, dispersion changes the $\alpha_{ti}(\vec{x})$ functions that appear in the partial derivative terms of the PDE, i.e.,

$$\alpha_{ti}(\vec{x}) = B_{ti}^*(\vec{x}) + \hat{b}_{ti}(\vec{x}) - D_{ti}^*(\vec{x}), \quad (14)$$

On the other hand, \vec{x} -groups vanish from the population at rate $\lambda_t(\vec{x})\theta_t(\vec{x})$ due to dispersion, which shows up on the right hand side of the PDE. Thus dispersion is different from the other events we have considered since it has effects at the individual level and group level.

4 Rate functions used in the examples of section 4

4.1 Example 4.1: Hunter-Gatherers

Example 4.1 in the main text studies the evolution of cooperation in a model of hunter-gatherer tribes, with properties,

1. Defectors have higher birth rates than cooperators in every tribe.
2. Larger tribes, and tribes with a larger proportion of defectors, are more likely to fission.
3. Smaller tribes, and tribes with a larger proportion of defectors, are more likely to die of extinction.

The birth rates for cooperators and defectors in an (x, y) -group have the form

$$B_c(x, y) = \beta + \frac{x}{x+y}b - c \quad B_d(x, y) = \beta + \frac{x}{x+y}b.$$

We assume the death rates are equal for cooperators and defectors, and proportional to the population size of the tribe, i.e., $D_c(x, y) = D_d(x, y) = D(x, y)$, where

$$D(x, y) = (x + y)\gamma.$$

In the main text, we study both the stochastic (simulation) and the deterministic (PDE) version of the model, and we assume the following parameter values: $\beta = 0.04$, $b = 0.05$, $c = 0.02$, and $\gamma = 0.0008$.

In a group of only cooperators, the birth and death rates are equal when $x = (\beta + b - c)/\gamma = 87.5$, and in a group of only defectors, the birth and death rates are equal when $y = \beta/\gamma = 50.0$. We assume that cooperators give birth to defectors and defectors give birth to cooperators 5% of the time, i.e., mutation rates are $\mu_{cd} = \mu_{dc} = .05$, so neither type can go permanently extinct.

The group-level events in the model are fissioning and extinction. The fissioning rate function is

$$f(x, y) = f_1x + f_2y,$$

and the extinction rate function is

$$e_t(x, y) = e_1G(t)e^{-(e_2x+e_3y)},$$

where $f_1 < f_2$ and $e_2 > e_3$, which is in the spirit of model assumptions. The extinction rate increases as the number of tribes in the population increases due to the $G(t)$ term, but the fissioning rate of a group only depends on the state of the group. The number of tribes in the population therefore cannot grow indefinitely. We set the parameters to $f_1 = .0008$, $f_2 = .0014$, $e_1 = .02$, $e_2 = .20$, $e_3 = .18$, and use the simple uniform fissioning density (5).

4.1.1 Stochastic simulation

An example of a stochastic simulation of the hunter-gatherer model can be found at the following link: <http://www.youtube.com/watch?v=EBW41K3xw0I&feature=youtu.be>.

4.1.2 Numerical solution of the PDE

The dynamical equation for the two-level population dynamics (equation (8) in the main text) is a first order hyperbolic PDE, and there are well established numerical techniques that can be used to solve them, see e.g. LeVeque (2005).

4.2 Example 4.2: Evolution of stickyness

Example 4.2 is a model of the evolution of simple multi-celled organisms in an environment containing predators, with the following properties,

1. Normal cells reproduce at a faster rate than sticky cells.
2. Stickier organisms are less likely to fission and more likely to fuse.
3. Smaller organisms are more likely to be eaten.

There are, of course, many ways to design a model with these properties. Our particular example works as follows. The population of predators satisfies

$$P'(t) = k_1F(t) - k_2P(t), \tag{15}$$

where $F(t)$ is the amount of “food” (small organisms) in the population at time t . We set

$$F(t) = \int \int \theta_t(x, y)e^{-k_3(x+y)} dy dx, \tag{16}$$

which incorporates the assumption that smaller organisms are more likely to become food. In our numerical example, the parameters are set to $k_1 = .05$, $k_2 = .001$, and $k_3 = .01$.

The population dynamics for sticky and normal cells within the organisms are specified by

$$\alpha_s(x,y) = \left(\beta_s \left(1 - \frac{x+y}{\gamma_1}\right)^+ - \delta_0 + \delta_1 \left(1 - \frac{x+y}{\gamma_1}\right)^- \right) x, \quad (17)$$

and

$$\alpha_n(x,y) = \left(\beta_n \left(1 - \frac{x+y}{\gamma_1}\right)^+ - \delta_0 - \delta_1 \left(1 - \frac{x+y}{\gamma_1}\right)^- \right) y, \quad (18)$$

where $\beta_s < \beta_n$, reflecting the assumption that normal cells divide faster than sticky cells. The rationale for (17) and (18) is that organisms are roughly spherical, and the surface cells of an organism must supply all its cells with food. Since the fraction of cells on the surface of a sphere decreases with the size of the sphere, there is a maximum organism size, γ , that can be supported: below this critical population level the cell birth rates are positive, and above it they are zero. Cell death rates have a base value, δ_0 plus an additional factor if the population of cells is above the critical size. In our numerical example, the parameters are set to $\beta_s = .060$, $\beta_n = .065$, $\gamma_1 = 800$, $\delta_0 = .001$, and $\delta_1 = .00001$.

We use the extinction rate function

$$e_t(x,y) = e_1 e^{-k_3(x+y)} P(t) + e_2 (N_s(t) + N_n(t)), \quad (19)$$

where $N_s(t)$ and $N_n(t)$ are the total populations of sticky cells and normal cells at time t (respectively) given by equation (3) in the main text, $k_3 = .01$ is the same parameter that shows up in (15), and $e_1 = .001$, $e_2 = 10^{-8}$. The first term on the right side of (19) therefore corresponds to extinction by predation, and the second term corresponds to extinction by starvation.

To reflect the assumption that stickier organisms are less likely to fission, our fissioning rate function has the form

$$f(x,y) = \frac{f_1 e^{f_2 y}}{1+x}, \quad (20)$$

with $f_1 = .003$ and $f_2 = .015$. We use the simple uniform fissioning density (5).

Finally, to reflect the assumption that stickier organisms are more likely to fuse, our fusing rate function has the form

$$\rho((x,y), (u,v)) = \rho_0 \frac{x}{x+y} \frac{u}{u+v} e^{-(x+y+u+v)/\gamma_2} \mathbf{1}_{(x+u < 1000 \ \& \ y+v < 1000)}, \quad (21)$$

where $\rho_0 = .025$ and $\gamma_2 = 200$. (The indicator function in (21) prevents organisms with more than 1000 normal or sticky cells from forming by fusion. Its only purpose is to make the numerical solution of the PDE more efficient, as it assures that $\theta_t(x,y)$ has support on a bounded region.)

5 References

1. LeVeque, R.J., (2005), "Numerical Methods for Conservation Laws", 2nd Ed., Birkhauser.
2. Okasha, S., (2006), *Evolution and the levels of selection*, Oxford University Press.
3. Puhalskii, A. and Simon, B., (2012), "Discrete Evolutionary Birth-Death Processes and their Large Population Limits", *Stochastic Models*, to appear.
4. Ross, S.M., (2012), *Simulation*, 5th edition, Academic Press.
5. Simon, B. (2010), "A Dynamical Model of Two-Level Selection", *Evolutionary Ecology Research*, 12. 555-588.
6. Simon, B., Fletcher, J.A., and Doebeli, M., (2011), "Hamilton's Rule in Multi-Level Selection Models", *Journal of Theoretical Biology*, to appear.