It takes grouping and cooperation to get sociality

Matthijs van Veelen [a,*], Julián García [a,b], Leticia Avilés [c]

[a] CREED, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
[b] Faculty of Economics and Business Administration, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
[c] Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

ABSTRACT

Cooperation and grouping are regularly studied as separate traits. The evolution of sociality however requires both that individuals get together in groups and that they cooperate within them. Because the level of cooperation can influence selection for group size, and vice versa, it is worth studying how these traits coevolve. Using a generally applicable two-trait optimization approach, we provide analytical solutions for three specific models. These solutions describe how cooperative associations of non-relatives evolve, and predict how large and how cooperative they will be. The analytical solutions help understand how changes in parameter values, such as the group carrying capacity and the costs of cooperation, affect group size and the level of cooperation in equilibrium. Although the analytical model makes a few simplifying assumptions—populations are assumed to be monomorphic for grouping as well as for cooperative tendencies, and group size is assumed to be deterministic—simulations show that its predictions are matched quite closely by results for settings where these assumptions do not hold.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The formation of groups and the evolution of cooperation within them—the two essential ingredients of sociality (Alexander, 1974)—have largely been treated separately in ecological and evolutionary models of social evolution. Group formation and group size evolution have been the main focus of ecological models such as for instance those of Giraldeau and Caraco (1993) and Higashi and Yamamura (1993). The evolution of cooperation, on the other hand, has been studied with models in which group size is fixed, or, more generally, with models in which the structure of who can affect whom by cooperating or not cooperating is exogenously given (see the overview by Lehmann and Keller, 2006, and references therein). It is natural to expect, however, that group size and cooperation will have an effect on each other. Whether or not selection will for instance favour an increase in cooperation may very well depend, not only on the current level of cooperation, but also on current group size. The same holds for grouping; whether or not it pays to form larger groups may depend on the level of cooperation as well as on the size of the group. In a series of recent papers Avilés (2002) and Avilés et al. (2002, 2004) use an individual-based and genetically explicit simulation model to explore the evolution of grouping

and cooperation, treating these two traits as coevolving dynamic variables. These simulations are used to describe how group size and cooperation respond jointly to ecological and demographic parameters for this particular model.

Using a two-trait optimization approach, here we provide analytical solutions for the joint evolution of cooperation and grouping, while allowing for the possibility that they feed back into each other. We consider three alternative models. The first is the model from Avilés (1999, 2002) and Avilés et al. (2002, 2004). The second and third model are alternative possibilities for how cooperation and grouping can affect fitness. We explore analytically the similarities and differences among the three models and consider the different ecological circumstances they may represent.

We also compare the analytical solutions to simulation results. The reason why these two might diverge is that the analytical approach makes the simplifying assumption that populations are monomorphic in equilibrium. In the simulation model of Avilés (2002) and Avilés et al. (2002), that is not the case; because grouping and cooperative tendencies are coded as polygenic traits, and because they evolve to intermediate levels, sexual reproduction maintains some variance in the population. In addition, group size is inherently stochastic in the simulation model, while in the analytical model we assume that group size is deterministic. The simulation results are nonetheless rather similar to the analytical predictions of the simplified monomorphic model. With the analytical solutions for the simplified case at hand, we are therefore much better able to describe and

understand how and why the equilibrium values for group size and cooperativeness react to changes in parameter values in the simulation model. We also show that the same holds for other models that are implemented in simulations in a similar way.

2. The model

We assume that individuals have the possibility to come together in groups in order to perform a task that might be done more efficiently together than alone. There are two things that will matter for group productivity, or the success with which the task is performed: group size and the level of cooperation. In the model, both of them ultimately depend on individual character-istics. Group size depends on how eager individuals are to get together and form groups. Once within a group, individuals can contribute to the overall success of the group by being cooperative. Cooperation increases total group productivity, but lowers the relative fitness of cooperators within their group. A natural setting one can think of is one where individuals form one-generation breeding associations, after which the offspring produced within the groups join a global pool from which they disperse to restart a new cycle of group formation.

Following Avilés (1999, 2002) and Avilés et al. (2002, 2004), we assume that the number of offspring produced by an individual is a function of the size of the group it is in, of the cooperativeness of the other group members, and of the cooperativeness of the individual itself. We also assume that this function is hump-shaped with respect to the size of the group. We will focus on three such functions. The first is taken from Avilés (1999, 2002), the other two are different specifications that imply different dynamic behaviour.

$$f_i(n,\overline{\gamma},\gamma_i) = e^r e^{-cn} n^{\overline{\gamma}}(1+\beta(\overline{\gamma}-\gamma_i)) \tag{1}$$

$$g_i(n,\overline{\gamma},\gamma_i) = e^r e^{-cn}(1+n\overline{\gamma})(1+\beta(\overline{\gamma}-\gamma_i)) \tag{2}$$

$$h_i(n,\overline{\gamma},\gamma_i) = e^r e^{-cn} n\overline{\gamma}(1+\beta(\overline{\gamma}-\gamma_i)) \tag{3}$$

In all three functions, n is the size of the group the individual is in, $\overline{\gamma}$ represents the average cooperative tendency of the members of the group, γ_i is the cooperative tendency of the individual itself, r is an intrinsic rate of growth parameter, c is the inverse of a group carrying capacity parameter, and β reflects the cost of coopera-tion. Given these definitions, e^r represents the intrinsic growth in the absence of cooperative or competitive interactions, e^{-cn} reflects the negative effects of crowding and competition, given limited resources available to a group, and $1+\beta(\overline{\gamma}-\gamma_i)$ describes how individuals that are more cooperative than average within their group, have a lower than average fitness compared to their fellow group members, and vice versa. The factor that differs between the three—$n^{\overline{\gamma}}$ in (1), $1+n\overline{\gamma}$ in (2) and $n\overline{\gamma}$ in (3)—represents the effect of cooperation on the average productivity within the group.

In Avilés et al. (2002, 2004) the first three factors are presented together, while the last one is presented separately. Here we present the compound function, because for the analytical derivations it is more convenient to have the whole formula at once. It is nonetheless worth realizing that the first three terms represent the average reproductive success of members of a group, or, when multiplied by the total number of group members, the total group productivity (i.e., the size of the pie), while the last term describes the share an individual gets of the group productivity pie. This helps interpreting the derivations made on the individual level, but it also helps understand how this is still a model where differential group productivity and two levels of selection are involved (see also the end of Section 3). We

also note that in this model all group members, including the cooperator, receive a share of the benefits of the cooperative behaviour, which makes it, in Pepper's (2000) terminology, a "whole group" rather than an "others only" model of social evolution.

For Eqs. (1) and (3) the functions are all one-humped with respect to n whenever $\overline{\gamma}$ is larger than zero; the function is only decreasing for $\overline{\gamma}=0$ with Eq. (1) and flat for $\overline{\gamma}=0$ with Eq. (3). For Eq. (2) the function is one-humped when $\overline{\gamma}>c$, and only decreasing for $\overline{\gamma}<c$.

3. Results

If we simplify matters by assuming that in equilibrium a population consists of individuals that all have equal grouping and cooperative tendencies, and that group size is a deterministic function of an individual's grouping tendency and of the average grouping tendency in the rest of the population, then we can think of evolution as a two-trait optimization process. Taking the derivative of an individual's fitness function with respect to its own cooperative tendency γ_i then indicates when cooperative tendencies are expected to go up and when they are expected to go down. Taking the derivative with respect to an individual's grouping tendency does the same for grouping tendencies. The derivations are provided in Appendix A, and the resulting conditions are summarized in Table 1 and Fig. 1.

The equations tell us that in order to know the direction of selection, we can just look at two values: group size and cooperativeness. The state space of the two traits is divided into four regimes by the conditions from Table 1, as depicted in Fig. 1 for each of the three functions. The intersection of the two lines is the point at which individual fitness is maximized with respect to cooperation and grouping tendencies and therefore a fixed point of the dynamics. In Appendix B we show that, under assumptions concerning the way in which individual grouping tendencies translate to group size, these fixed points are stable, so in the absence of noise, we can expect the population to settle there. With Eq. (1) trajectories spiral towards the fixed point. With Eq. (2) it depends on the relative speed of selection for both traits (see Appendix B) whether or not the trajectories spiral. With Eq. (3) one can directly see from Fig. 1c that spiralling is not possible; because the isocline for group size is vertical, trajec-tories cannot cross it.

The simple, first prediction that follows from the analysis above is that a population will converge to the intersection of the two lines in the pictures. Levels of cooperation and group sizes at equilibrium thus depend only on the group carrying capacity c and the costs of cooperation parameter β, and not on the intrinsic rate of growth r. The only effect r can have—everything else being equal—is that if it falls below a threshold level, the population will go extinct. A change in r however has no effect on predicted group size and level of cooperation.

Table 1
Conditions under which group size (derivative to individual grouping tendency is 0) and level of cooperation (derivative to individual cooperation tendency is 0) are expected not to change.

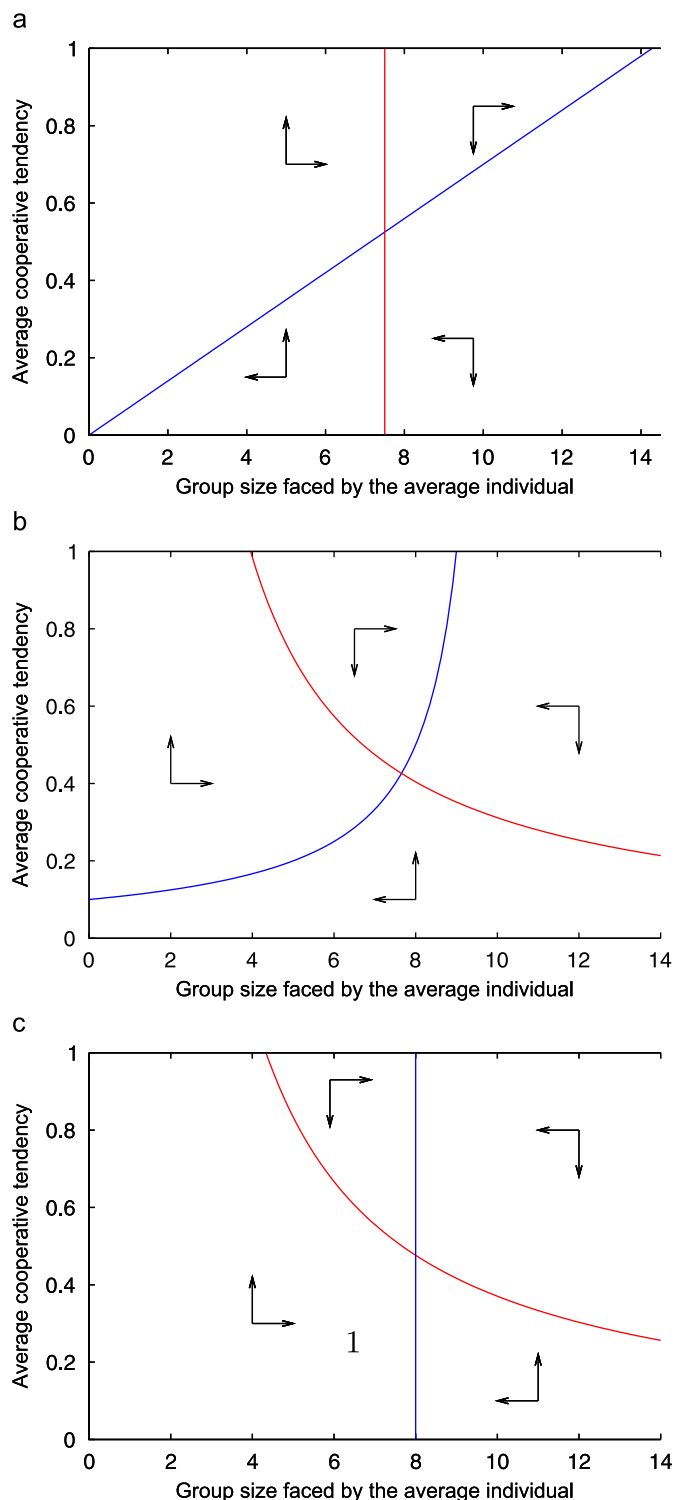| Fitness function	Derivative to grouping is 0	Derivative to cooperation is 0
f_i	$n=\dfrac{\overline{\gamma}}{c}$	$\beta=\dfrac{\ln n}{n-1}$
g_i	$n=\dfrac{1}{c}-\dfrac{1}{\overline{\gamma}}$	$\overline{\gamma}=\dfrac{1}{\beta(n-1)}-\dfrac{1}{n}$
h_i	$n=\dfrac{1}{c}$	$\overline{\gamma}=\dfrac{1}{\beta(n-1)}$

Fig. 1. Direction of selection for the three functions. The red lines separate the parameter values for which the derivative of the functions f_i, g_i and h_i (Eqs. (1)–(3)) to cooperation is positive (left of it) from those for which it is negative (right). Similarly, the blue lines separate parameter values for which the derivative to grouping is positive (left) and negative (to the right).

and an increased group size in (1) and (2), while group size remains constant in (3). If group carrying capacity goes up (that is, if c goes down) then the group size isocline (blue in Fig. 1) rotates clockwise, with the origin fixed, in Fig. 1a, and moves to the right in Fig. 1b and c. Group size then goes up in models (2) and (3) but remains constant in (1), while cooperation goes down in all models. The effect of changes in group carrying capacity on group size in model (1) can at first be perceived as a bit counterintuitive; one would expect that group carrying capacity will first of all have an effect on group size, and perhaps also on cooperation. Here, however, the prediction is that an increase of the group carrying capacity only affects cooperation (it goes down) while group size remains the same. This can be understood, if we see that after an increase in group carrying capacity, initially groups will indeed grow in size. This however then makes cooperation unfavourable, which reduces cooperativeness, and that, in turn, causes large groups to be selected against. The population will then spiral towards the new equilibrium, which has the same group size as before the increase in group carrying capacity, but a lower level of cooperation (Fig. 2).

It is important to realize that this two-trait optimization approach is not at all at odds with this being a model with different levels of selection, when cooperation is concerned
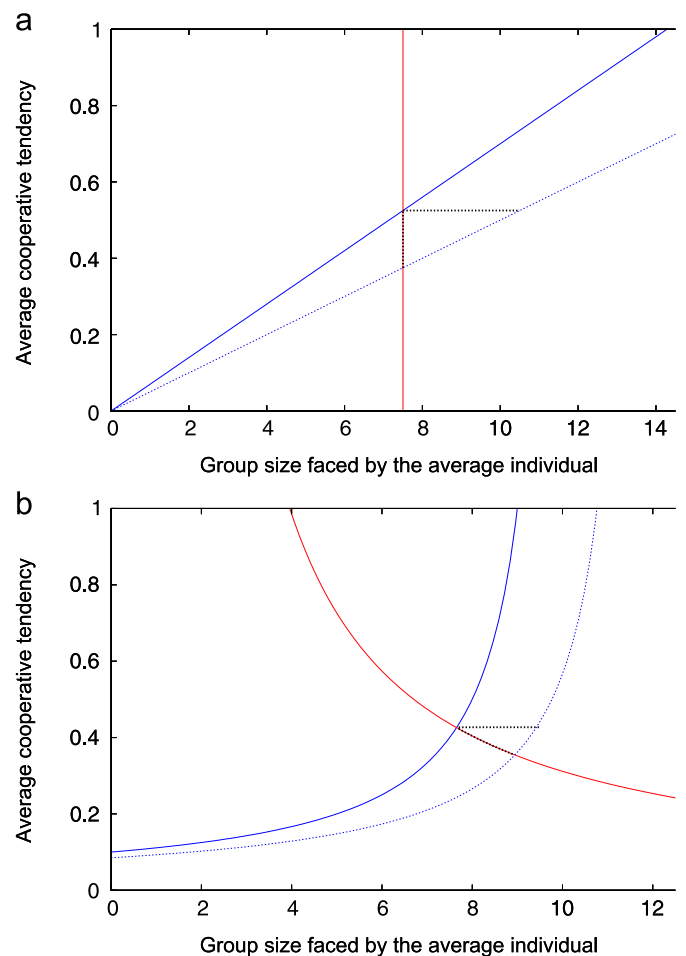

Fig. 2. Increase of group carrying capacity. In absence of a feedback effect through the level of cooperation, one would expect that an increase in group carrying capacity would make groups grow larger. A shift of the isocline for Eq. (1), that reflects an increase in group carrying capacity, however leads to a shift of the stable fixed point to a new equilibrium with the exact same group size, but a lower level of cooperation (a). The effect of the feedback is smaller for Eq. (2) than it is for Eq. (1); there we do predict an increase in group size, although smaller than it would have been without feedback (b).

We also see how changes in values of the parameters c and β affect the stable fixed point. If current cooperation is strictly between 0 and 1, then a decrease in β, that is, a decrease in relative fitness costs of cooperation, will make the cooperation isoclines (red in Fig. 1) shift to the right. This implies that the new fixed point has an increased level of cooperation for all models,

(see also van Veelen, 2009). Obviously, within groups, individuals with a lower than average cooperative tendency do better than individuals with a higher than average cooperative tendency. Also, between groups, groups with a higher average cooperative tendency do better than groups with a lower average cooperative tendency. So within groups, cooperativeness is selected against, and between groups cooperativeness is selected for. With cooperative tendencies being randomly distributed over groups, we know that one can determine the direction of selection by considering how changes in the individual's own behaviour would affect its own absolute fitness. Even if less cooperative individuals do better than more cooperative individuals within the group, it can very well be that changing from cooperative behaviour to less (or un-) cooperative behaviour actually harms the individual's own absolute fitness (see for instance Kerr and Godfrey-Smith, 2002, or the instructive table in Wilson, 2004). Taking the derivative to one's own cooperative tendency and setting it to 0 does apply this criterion; it looks at the effect of an in- or decrease of the individual's own cooperative tendency on its own fitness. As we will see below, it is an approximation; by taking derivatives for a monomorphic population, it ignores the possibility that this effect may vary according to the composition of the rest of the group. Below we will also see that this is nonetheless a good approximation (see also Appendices A and C).

4. Match between model and simulations

In order to determine to what extent the analytical predictions are matched by simulation output, we coded a new Java version of the Avilés et al. (2002) and Avilés (2002) simulation model and extended it so that it contains all three functions (Eqs. (1)–(3)) and a few additional options. The original simulation model is described in Avilés et al. (2002, 2004) and Avilés (2002). A complete and detailed description of the new version of the model, with extensions and modifications, along with an on-line version of the program, is provided at http://staff.feweb.vu.nl/j.garcia/sociality.

Fig. 3 shows that the match between the analytical predictions and the simulation results is rather close. There are some differences though; the simulation results are slightly shifted towards the right for all three models (that is, we find greater group sizes than predicted; see also Fig. 4). We also find some distortion towards the edges, in particular where cooperation is close to 0 or close to 1. Before describing what causes these differences, it is worth noting that in the figures, we plot the group size faced by the average individual, rather than the average group size. The group size faced by the average individual is what Jarman (1974) called the "typical group size" (see also Jarman and Jarman, 1979; Jarman, 1982; Reiczigel et al., 2008). In order to see why this is the appropriate measure, we should realize that in the simulations, an individual is characterized, not only by a cooperative tendency, but also by a grouping tendency. This grouping tendency influences the expected size of the group it ends up in; a higher grouping tendency makes the group an individual is in accept more group members, and hence grow larger, while a lower grouping tendency makes the group it is in accept less, all in expectations. Grouping tendencies therefore should rise as long as it would be advantageous for individual group members to have a higher expected group size. The value that matters therefore is the group size that an individual can expect to face. This is not the same as the average group size; these two quantities only coincide if all groups are of equal size, but if groups differ in size, the former is larger than the latter, because large groups not only are large, but they also account for more individuals that face a large group.

4.1. Why the simulation outcomes are shifted

The shift of the simulation outcomes to the right, relative to the analytical predictions (Figs. 3 and 4), reflects the asymmetry of the functions around the optimal values of n and γ, combined with the fact that the analytical approach makes the simplifying assumption that at equilibrium, populations are composed of equally sized groups and are monomorphic for the cooperative tendency of individuals. In the simulations, in contrast, cooperative and grouping tendencies are coded as diploid polygenic traits, and because they evolve to intermediate levels, sexual reproduction maintains some variance in the population. Also the group formation process implies that the group size an individual will face is not deterministic. Even if the population were monomorphous both for grouping and for cooperative tendencies, the group size that an individual faces would still be a draw from a random distribution over group sizes. The size and cooperativeness of the group an individual will find itself in therefore is a random variable that depends on an individual's own cooperative tendency and its grouping tendency, as well as on those of the others in the population. Below and in the Appendix, we focus on group size as a variable.

The asymmetry of the fitness function around the optimal values of n here means that being smaller than the optimal group size is worse (further away from the optimal fitness) than being just as much larger. Since individuals cannot choose a fixed group size, but are restricted in their choice between random distributions over group sizes, this implies that the isoclines of the true model can differ from the isoclines from the monomorphic model. How much they will differ, depends on the shape of the random distributions of group sizes the individual can choose from, and the shape of the fitness function. Because here, being smaller than optimal is worse than being larger than optimal, the expected size that comes with optimizing over the whole distribution is a bit larger than the optimal size in the monomorphic model, which implies that the group size faced by the average individual in equilibrium moves a bit to the right. This equilibrium would in fact most accurately be described as a fixed point in distributions. Appendix C.1 gives examples that illustrate how the group size faced by the average individual in the equilibrium distribution can be larger or smaller than the fixed point of the monomorphic model, due to the asymmetry in the fitness function and the set of probability distributions that an individual can choose from. Appendix C.2 describes how the population being a balanced polymorphism amplifies the effect of the asymmetry in the fitness function. For grouping tendency, and with sexual reproduction, one could therefore say that the asymmetry of the fitness function affects the mean twice; once because there is uncertainty about the group size one will find itself in, given ones own grouping tendency, and once because there is uncertainty about the grouping tendency of one's offspring. In both cases it is better to be safe than sorry; ending up in too small a group oneself is relatively bad—compared to ending up in a too large group—and having offspring that happens to have a lower grouping tendency than oneself is worse than offspring that happens to have a higher grouping tendency. Both uncertainties thereby increases the mean, because deviations from the maximum of the fitness function to smaller groups are worse than deviations to larger groups.

Although the variance in cooperativeness is lower than the variance in group size, a similar effect occurs for cooperativeness. Fig. 4 has, for all three functions, singled out one of each of the isoclines, and compared the prediction for the monomorphic model with the averaged outcome of the simulation model.
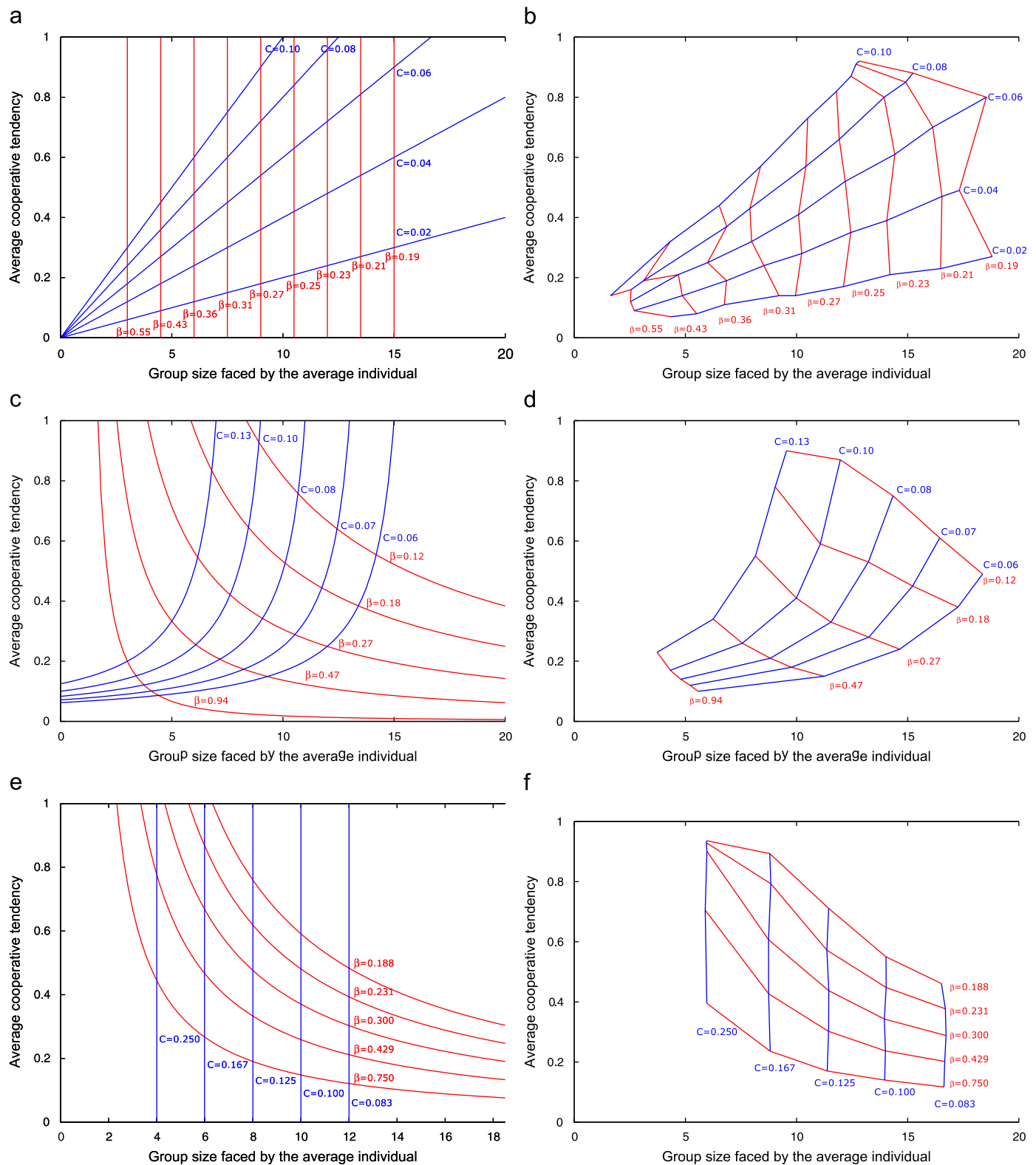
Fig. 3. Analytical predictions and simulation results. The figures on the left depict the analytical predictions for the levels of cooperation and for group sizes for, from top to bottom, the functions (1)–(3). The intersections of the isoclines for the β's (red) and the c's (blue) are the fixed points of the dynamics for monomorphic populations for different values of β and c. The results of the simulations are pictured in the figures on the right. Runs, two for every combination of parameter values, lasted 15 000 generations, of which we took the last 6000–12 000 to compute the average cooperative tendency and group size, depending on the time it takes for runs to get close to a stable fixed state. With the last function, initiation at a level of cooperation of 0 leads to immediate extinction of the population, because $h_i(n,0,0)=0$. The fixed points are therefore only reached if the population is initiated at a sufficiently large level of cooperativeness. This would match the situation of a habitat in which sociality could not evolve from solitary living, but where already social animals could invade.
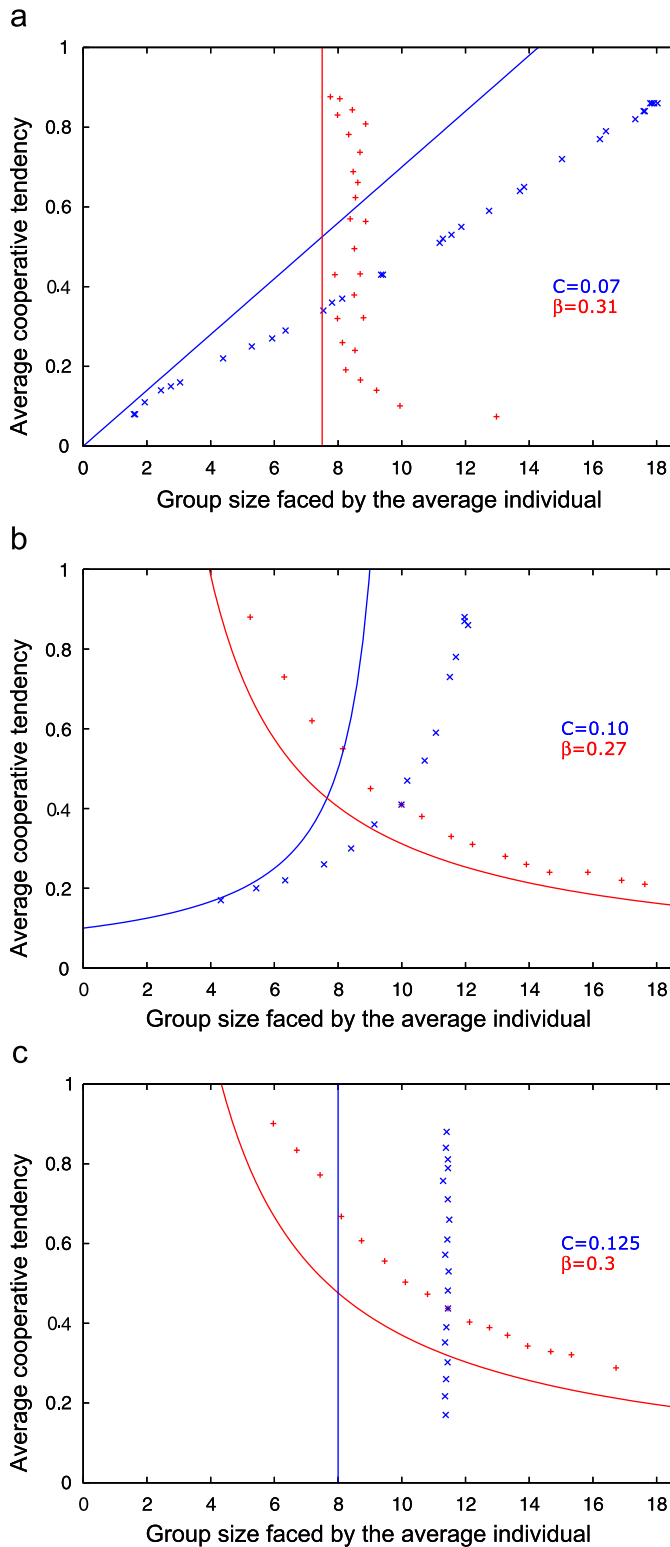
Fig. 4. Shifted isoclines for functions f_i, g_i and h_i (Eqs. (1)–(3)). The isoclines from the simulation model are estimated by keeping the β constant and varying the c (red) and vice versa (blue). Apart from the shifts caused by the group size being a distribution and the functions f_i, g_i and h_i not being symmetric, we also see that closer to the sides, the red isocline is curved.

4.2. Distortion of the cooperation isocline towards the edges

In Fig. 4, we see that the cooperation isocline is distorted towards the edges. That is especially visible in Fig. 4a; close to the edge where cooperation is 0, the isocline from the simulations

bends to the right, and close to the edge where cooperation is 1, it bends to the left. These bends are caused by the fact that close to the edges, mutations become biased against the common allele. In the simulation program, inversions and translocations take precedence over point mutations. This implies that for values of the average cooperative tendency that are not close to 0 or 1, the most common mutations are inversion and translocation, which are unbiased. The whole mutation process therefore is almost unbiased, when not close to the edges. Towards the edges, however, mutations become more likely to actually change the phenotypic value of individuals; if it is not possible to change the genome by inversion or translocation (i.e., if the individual has phenotypic value 0 or 1), then a point mutation will occur. Since point mutations are biased against the common allele, this implies that the mutation process also becomes increasingly biased against the common allele towards the edges. A precise description of the mutation process is provided with the online version of the simulation program, and helps understand this effect better.

5. Discussion of the results

Despite being essential ingredients of sociality (Alexander, 1974), grouping and cooperation have mostly been considered in isolation (see Avilés, 2002 and references therein). Using a two-trait optimization approach, we have derived the first analytical solutions for the levels of cooperation and group sizes that are expected to arise from the joint evolution of these two traits. The analysis shows that the tendency to form groups and the tendency to cooperate interact, and that the shape of the functions that describe the number of offspring of an individual makes a difference for how changes in parameters change equilibrium outcomes. This indicates the importance of not looking at the two traits in isolation; allowing them to evolve together can give predictions that differ from what one would predict if only one of the two is allowed to evolve.

Because not all functions give the same results, and because we want to be able to match different functions with different biological situations, it is worth figuring out what characteristics they share, and what sets them apart. The first characteristic they share is that, for constant levels of cooperation, they all are hump-shaped with respect to group size (or more precisely: they do not have more than one local maximum). This can be regarded as a rather reasonable assumption, reflecting the idea that there is an optimal group size for the performance of a task. A humped-shaped function has in fact been demonstrated for a variety of social organisms (e.g., Caraco and Wolf, 1975; Nudds, 1978; Buss, 1981; Raffa and Berryman, 1987; Heinsohn, 1992; Cash et al., 1993; Komdeur, 1994; Wiklund and Andersson, 1994; Booth, 1995; Jeanne and Nordheim, 1996; Avilés and Tufiño, 1998) and is a common assumption for models of group living (Vehrencamp, 1983; Pulliam and Caraco, 1984; Slobodchikoff, 1984; Giraldeau, 1988; Emlen, 1991; Krause and Ruxton, 2002).

The second common characteristic is that for small groups, it pays off for an individual to cooperate (the derivative of f_i to γ_i is larger than 0), while cooperating becomes unfavourable if an individual finds itself in a large group. What "large" is, that is, from where onwards groups are so large that an increase in ones cooperation level decreases ones own fitness, of course depends on parameter values and on current cooperativeness, but the idea is that there is such a point, in the same way as that there is an optimal group size. Again, this can be seen as reasonable, if the function reflects the success with which a task is performed; if groups are very large, cooperating—or cooperating more—may pay off too little to compensate for the costs.
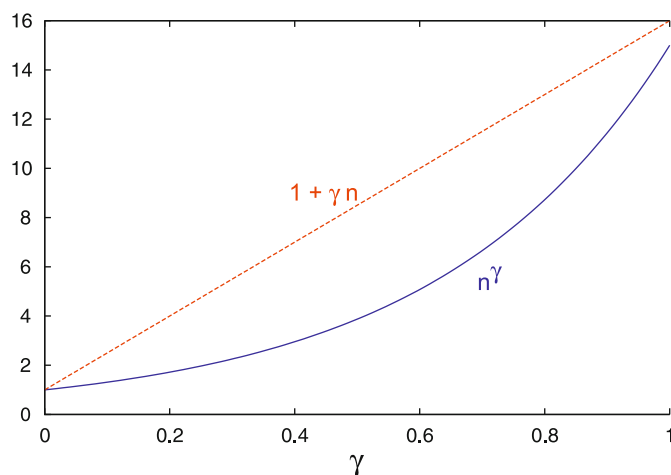
Fig. 5. The terms $n^{\overline{\gamma}}$ and $1+n\overline{\gamma}$ as functions of overall cooperation level $\overline{\gamma}$ with $n=15$.

If we indeed have two such lines, one representing where the maximum of the function with respect to group size lies, and one representing where cooperation stops being to the individual's own benefit, then the point where the lines intersect constitutes a fixed point of the dynamics. (There can also be fixed point where the lines meet the boundary; these fixed points must even be there if the two lines for instance do not cross.)

The difference between the functions lies in how extra overall cooperativeness translates to higher values of the three functions. For that purpose, we only need to look at the part that is different between them, that is, $n^{\overline{\gamma}}$ in Eq. (1), $1+n\overline{\gamma}$ in (2) and $n\overline{\gamma}$ in (3). With Eq. (1), an increase in overall cooperation level makes more of a difference if $\overline{\gamma}$ is close to 1 than if it is close to 0 (see Fig. 5). In Eqs. (2) and (3), an increase in overall cooperation level makes the same difference everywhere. The term $n\overline{\gamma}$ in these two functions can therefore be interpreted as the total amount of cooperation; it is simply the sum of the cooperative tendencies of all individuals. A model where every extra bit of help adds just as much would be appropriate for situations where, for instance, resources of an ephemeral and unpredictable nature need to be located, so that every additional participant adds just as much to how much of the resource will be found (e.g., bee hives or marine colonial birds searching for flowers or for moving schools of fish, respectively). Looking at those terms as a function of n can also be instructive; in Eq. (1) there are diminishing returns to adding equally cooperative individuals, which would apply to situations such as the joint capture of large prey or the warning of a predator's arrival, where every additional participant would have a smaller contribution to the success of the task.

One can also think of other functions that may describe how cooperation translates into more offspring. One could, for instance, take the mirror image of (1), where the effect of an overall increase of cooperation (not to be confused with the effect of an increase in an individual's level of cooperation) levels off when cooperativeness goes up, or a function that is S-shaped. The latter would be appropriate for situations when there are thresholds involved in the efficient performance of a task, such as tree killing bark beetles that need to overcome the defenses of live trees, which they do by attacking in mass (Raffa and Berryman, 1987). For those functions, obviously the analytical approach from this paper will again provide predictions for equilibria.

Furthermore, one more detail of the functions may be interesting for their interpretation, and that is that $n^{\overline{\gamma}}=1$ for $n=1$, regardless of $\overline{\gamma}$. This makes sense, because cooperation would be meaningless if there is no-one to cooperate with. On the

other hand, $1+n\overline{\gamma}$ increases with $\overline{\gamma}$, also if $n=1$. This can make sense for cases where cooperation can be interpreted as the provision of a public good, such as building a (possibly communal) web, building protective structures for a (possibly communal) nest, detoxifying waste products from the common environment or cleaning up the nest. Also on one's own, spending some effort on building a web or on nest defense can be sensible, but it becomes a public good, or a collective effort, if others join the group.[1]

Eq. (3), but not Eq. (2), on the other hand, has the peculiarity that if average cooperativeness $\overline{\gamma}$ is 0 in the latter, then $n\overline{\gamma}$ is also 0, which implies that the fitness of all individuals is 0. This may describe a habitat in which a certain species could not live without cooperating, whatever its group size, including size 1. Such a harsh environment could, on the other hand, be colonized by an already social species that has evolved sufficiently high levels of grouping and cooperation in another habitat. The same applies to Eqs. (1) and (2) for values of r that are too low to have replacement in the absence of cooperation and grouping.

One can also make a mix of the functions (2) and (3) by replacing $1+n\overline{\gamma}$, or $n\overline{\gamma}$, by $\alpha+n\overline{\gamma}$. This new parameter α would then reflect the relative importance of the task that is performed together; the lower α, the more important the collective task. Analysis of the dynamics for this slightly more general function is not more complicated, and is done in Appendix A.

One point worth mentioning here is that the model from Avilés (2002) is classified in Lehmann and Keller (2006) as a model where altruism evolves due to a greenbeard effect. With the analysis in this paper, we can conclude that in their classification, it would fall under cooperation rather than under altruism, and would be explained by direct benefits instead of a greenbeard effect. More precisely, one can say that the dynamics in the model bring the population to a point at which the costs (c in Lehmann and Keller, 2006) of *marginal* changes in behaviour are 0; in equilibrium, derivatives to group size and, more importantly, to cooperation are 0. Alternatively, one can use the terminology of Wilson (1979), in which case that would translate to the dynamics bringing the population up to a point where further, *marginal* changes in behaviour are on the border between weak and strong altruism, or, in the terminology of Kerr and Godfrey-Smith (2002), between class I and class II altruism.

Finally we would like to point out that the ecology can put a limit on the growth of the overall population in different ways. One can assume that there is only a limited number of nesting sites, and in Avilés et al. (2002, 2004) this is implemented by assuming that the flow from the global pool towards the groups stops as soon as the last nesting site has been occupied by one individual. One can also assume that the global density negatively affects individual fitnesses. A third option is to assume that every generation, the global pool is reduced to a fixed number of individuals before they leave in order to form groups. For the simulations reported here, we chose the limited nesting sites model, but all three possibilities are all coded in the program. This program can be found online, and comes together with a precise description of the dynamics.

The most important distinctive feature of the global density dependence version is that this version, in combination with particular choices of the parameters, can allow for cyclical or chaotic behaviour of the size of the total population. This can

---

[1] A detail of lesser importance is that for $n^{\overline{\gamma}}$, the function f_i (Eq. (1)) increases if a single individual, regardless of its own cooperative tendency, is joined by another individual, even if the joining individual has cooperativeness $\gamma=0$, since $2^{\gamma/2}=(\sqrt{2})^{\gamma}>1^{\gamma}$. This peculiarity does not apply to groups with more than one initial individual, since $(n+1)^{n\gamma/n+1}=((n+1)^{n/n+1})^{\gamma}<n^{\gamma}$ if $n>1$. It is also excluded for all group sizes in Eqs. (2) and (3).

describe population dynamics in for instance social spiders (Avilés, 1997, 1999) and tree killing bark beetles (Berryman, 1973; Avilés, 2002). With cycles or chaotic behaviour at the population level, the optimum levels of grouping and cooperative tendencies then become moving targets. Nonetheless, average cooperative tendencies and group sizes are not that different from those in the limited nesting sites model. Compared to the theoretical prediction, the shift of the c-isocline—that reflects the group size dynamics—to the right is only a bit more pronounced if there are cycles or if there is chaotic behaviour. The reason is that the asymmetry of the fitness function around the optimum now not only punishes chance deviations from the optimal group size asymmetrically, but it also punishes deviations from the (moving) optimum over time asymmetrically (see also the online simulation model and results there). This also explains why r does have a small effect with global density dependence; by increasing the r we go from a stable system to cycles to chaos.

Acknowledgements

Appendix A. Derivations of the stability conditions

A.1. Function (1)

Before taking derivatives of the function f_i, we rewrite it, using

$$\overline{\gamma} = \sum_{j=1}^{n} \frac{\gamma_j}{n}$$

$$f_i(n,\overline{\gamma},\gamma_i) = e^r e^{-cn} n^{\overline{\gamma}} (1+\beta(\overline{\gamma}-\gamma_i))$$

$$= e^r e^{-cn} n^{(\sum_{j\neq i}^{n} \gamma_j/n + \gamma_i/n)} \left(1+\beta\left(\sum_{j\neq i}^{n} \frac{\gamma_j}{n} - \frac{n-1}{n}\gamma_i\right)\right)$$

Then we take the derivative to an individual's own cooperativeness:

$$\frac{df_i}{d\gamma_i} = e^r e^{-cn} n^{\sum_{j\neq i}^{n} \gamma_j/n} \frac{\ln n}{n} n^{\gamma_i/n} (1+\beta(\overline{\gamma}-\gamma_i)) + e^r e^{-cn} n^{\overline{\gamma}} \left(\beta\left(-\frac{n-1}{n}\right)\right)$$

$$= e^r e^{-cn} n^{\overline{\gamma}} \left(\frac{\ln n}{n}(1+\beta(\overline{\gamma}-\gamma_i)) - \beta\frac{n-1}{n}\right)$$

The first part of the fitness function—$e^r e^{-cn} n^{\overline{\gamma}}$—reflects group productivity, or the size of the pie, while the last term—$(1+\beta(\overline{\gamma}-\gamma_i))$—reflects the share that individual i gets. The first part of this derivative—$e^r e^{-cn} n^{\overline{\gamma}}(\ln n/n)(1+\beta(\overline{\gamma}-\gamma_i))$—therefore reflects the fact that if i increases its cooperative tendency γ_i, then the pie gets larger. The second part of this derivative—$e^r e^{-cn} n^{\overline{\gamma}}(-\beta(n-1)/n)$—reflects the fact that increasing γ_i also implies that i gets a reduced share of the pie.

Now we can evaluate this derivative in $\overline{\gamma}=\gamma_i$, which implies that $df_i/d\gamma_i=0$ if $\ln n=\beta(n-1)$. So if n^* solves $\beta=\ln n/(n-1)$, then any value for $\overline{\gamma}$ could be a fixed point with respect to cooperativeness, since the derivative $df_i/d\gamma_i$ goes through 0 at $\gamma_i=\overline{\gamma}$. If, however, $\beta<\ln n/(n-1)$, then $df_i/d\gamma_i$ is positive at $\gamma_i=\overline{\gamma}$, which implies that the average will be pulled up, whatever the average is, and if $\beta>\ln n/(n-1)$, then $df_i/d\gamma_i$ is negative at $\gamma_i=\overline{\gamma}$, and the average will be pulled down, whatever the average is.

Before taking the derivative to individual grouping tendency, we need to write the size of the group an individual will end up in as a function of its own grouping tendency and of the average grouping tendency in the rest of the population. In the simulations, the size of the group an individual will find itself in is a complex random variable that we cannot compute. It is however clear that this distribution depends on the grouping tendencies of all other individuals, on the grouping tendency of the individual itself, and on of the population size. The latter matters, because if the population is larger, then that means that there are more individuals soliciting for admittance, so the expected group size is also larger (see details in http://staff.feweb.vu.nl/j.garcia/sociality). Fortunately we will see that the equilibrium is invariant to what can be seen as mere rescaling of the relation between grouping tendency and group size, so this will turn out not to matter.

We simplify the stochastic simulation model to a deterministic version, where the group size an individual will find itself in depends on the average grouping tendency of the population, excluding the individual itself, and on the grouping tendency of the individual itself. The dependence on the population size will be suppressed, so at the onset, we will just be computing the derivative for a specific population size. We will, as mentioned above, find out that the specifics of the function will not matter, as long as it satisfies a very modest and natural assumption. Population size will therefore not feature in the final characterization of the equilibrium. This minimal assumption that we need is that the derivative of the group size an individual finds itself in to the grouping tendency of that individual itself is positive. (Although it is not used, it also seems obvious that the derivative to the average grouping tendency of the rest of the population should be positive.) So if δ_i is the individual's own grouping tendency, and $\overline{\delta}$ is the average grouping tendency in the population, then we assume that $\partial n(\overline{\delta},\delta_i)/\partial\overline{\delta}>0$ and $\partial n(\overline{\delta},\delta_i)/\partial\delta_i>0$.

Then we take the derivative to an individual's own grouping tendency as follows:

$$\frac{df_i}{d\delta_i} = \frac{\partial n}{\partial\delta_i}\{-ce^r e^{-cn} n^{\overline{\gamma}}(1+\beta(\overline{\gamma}-\gamma_i))\} + \frac{\partial n}{\partial\delta_i}\{\overline{\gamma}e^r e^{-cn} n^{\overline{\gamma}-1}(1+\beta(\overline{\gamma}-\gamma_i))\}$$

$$= \frac{\partial n}{\partial\delta_i} e^r e^{-cn} n^{\overline{\gamma}} \left(\frac{\overline{\gamma}}{n}-c\right)(1+\beta(\overline{\gamma}-\gamma_i))$$

At $\overline{\gamma}=\gamma_i$, we get

$$\frac{df_i}{d\delta_i} = \frac{\partial n}{\partial\delta_i} e^r e^{-cn} n^{\overline{\gamma}} \left(\frac{\overline{\gamma}}{n}-c\right)$$

Given that $\partial n/\partial\delta_i>0$, this implies that $df_i/d\delta_i>0$ if and only if $\overline{\gamma}/n>c$, that is, if $n<\overline{\gamma}/c$.

We therefore have a fixed point of the function f_i if

$$\beta=\frac{\ln n}{n-1} \quad\text{and}\quad n=\frac{\overline{\gamma}}{c}$$

A.2. Functions (2) and (3)

For the other two functions, we do the derivation by considering a function

$$g_{\alpha,i}(n,\overline{\gamma},\gamma_i) = e^r e^{-cn}(\alpha+n\overline{\gamma})(1+\beta(\overline{\gamma}-\gamma_i))$$

Before taking derivatives, we rewrite this function, using $\overline{\gamma}=\sum_{j=1}^{n}\gamma_j/n$

$$g_{\alpha,i}(n,\overline{\gamma},\gamma_i) = e^r e^{-cn} \left(\alpha+\sum_{j\neq i}^{n}\gamma_j+\gamma_i\right)\left(1+\beta\left(\sum_{j\neq 1}^{n}\frac{\gamma_j}{n}-\frac{n-1}{n}\gamma_i\right)\right)$$

Then we take the derivative to an individual's own cooperativeness:

$$\frac{dg_{\alpha,i}}{d\gamma_i} = e^r e^{-cn}(1+\beta(\overline{\gamma}-\gamma_i)) + e^r e^{-cn}(\alpha+n\overline{\gamma})\left(\beta\left(-\frac{n-1}{n}\right)\right)$$

$$= e^r e^{-cn}\left(1 + \beta(\overline{\gamma} - \gamma_i) - \beta\left(\frac{\alpha}{n} + \overline{\gamma}\right)(n-1)\right)$$

Now we can evaluate this derivative in $\gamma_i = \overline{\gamma}$, which implies that $dg_{\alpha,i}/d\gamma_i = 0$ if $1 = \beta(n-1)(\alpha/n + \overline{\gamma})$. Therefore, if $\overline{\gamma} = 1/\beta(n-1) - \alpha/n$, then cooperativeness does not change at $\overline{\gamma}$. If, however, $\overline{\gamma} < 1/\beta(n-1) - \alpha/n$, then $dg_{\alpha,i}/d\gamma_i$ is positive at $\gamma_i = \overline{\gamma}$, and the average will be pulled up until $\overline{\gamma} = 1/\beta(n-1) - \alpha/n$, and if $\overline{\gamma} > 1/\beta(n-1) - \alpha/n$, then $dg_{\alpha,i}/d\gamma_i$ is negative at $\gamma_i = \overline{\gamma}$, and the average will be pulled down until $\overline{\gamma} = 1/\beta(n-1) - \alpha/n$.

If we take the derivative with respect an individual's own grouping tendency, assuming that $\overline{\gamma} = \gamma_i$, we get

$$\frac{dg_{\alpha,i}}{d\delta_i} = \frac{\partial n}{\partial \delta_i}\{-ce^r e^{-cn}(\alpha + n\overline{\gamma})(1 + \beta(\overline{\gamma} - \gamma_i))\} + \frac{\partial n}{\partial \delta_i}e^r e^{-cn}\overline{\gamma}(1 + \beta(\overline{\gamma} - \gamma_i))$$

$$= \frac{\partial n}{\partial \delta_i}e^r e^{-cn}(\overline{\gamma} - c(\alpha + n\overline{\gamma}))$$

Hence, since $\partial n/\partial \delta_i > 0$, we find that $dg_{\alpha,i}/d\delta_i > 0$ if $\overline{\gamma} - c(\alpha + n\overline{\gamma}) > 0$, that is, if $n < 1/c - \alpha/\overline{\gamma}$.

We therefore have a fixed point of the function $g_{\alpha,i}$ if

$$\overline{\gamma} = \frac{1}{\beta(n-1)} - \frac{\alpha}{n} \quad \text{and} \quad n = \frac{1}{c} - \frac{\alpha}{\overline{\gamma}}$$

For the function g_i, we would take $\alpha = 1$, which means that the fixed point is described by

$$\overline{\gamma} = \frac{1}{\beta(n-1)} - \frac{1}{n} \quad \text{and} \quad n = \frac{1}{c} - \frac{1}{\overline{\gamma}}$$

For the function h_i, we would take $\alpha = 0$, which means that the fixed point is described by

$$\overline{\gamma} = \frac{1}{\beta(n-1)} \quad \text{and} \quad n = \frac{1}{c}$$

Appendix B. Stability of the fixed points

If we assume that the speed of selection is proportional to the derivatives computed above for the function f_i, then we get, after dividing both by $e^r e^{-cn}n^{(\overline{\gamma}-1)}$ and taking $\gamma_i = \overline{\gamma}$ and $\dot{n} = (\partial n/\partial \delta)\dot{\delta}$, the following differential equations:

$$\dot{n} = \left(\frac{\partial n}{\partial \delta}\right)^2 (\gamma - cn)$$

$$\dot{\gamma} = (\ln n - \beta(n-1))$$

Without specifying $\partial n/\partial \delta$ there is not much one can say, but if we assume that it is constant, and if we add an extra parameter a, which governs the relative speed between the two traits (that can be tuned by changing relative genome sizes in the simulations), then this is equivalent to looking at

$$\dot{n} = \gamma - cn$$

$$\dot{\gamma} = a(\ln n - \beta(n-1))$$

If we write that relative to the fixed point (n^*, γ^*), which then becomes the new origin, where n^* solves $\beta = \ln n/(n-1)$ and $\gamma^* = cn^*$, then we get

$$\dot{N} = \Gamma - cN$$

$$\dot{\Gamma} = a(\ln(N + n^*) - \beta(N + n^* - 1))$$

The matrix of first derivatives, evaluated in the origin, then becomes

$$\begin{bmatrix} 1 & -c \\ 0 & a\left(\frac{1}{n^*} - \beta\right) \end{bmatrix}$$

the eigenvalues of which are

$$\begin{cases} -\frac{1}{2}c \pm \frac{1}{2}\sqrt{c^2 - 4a} & \text{if } c^2 - 4a > 0 \\ -\frac{1}{2}c \pm \frac{1}{2}i\sqrt{4a - c^2} & \text{if } c^2 - 4a < 0 \end{cases}$$

Because both a and c are larger than 0, the eigenvalues, or their real parts, are smaller than 0, hence the fixed point is locally stable.

For the functions g_i and h_i a similar exercise can be done, but for the function h_i inspection of Fig. 1c is already enough; the isocline for group size is vertical there, which precludes trajectories crossing it. Hence all trajectories stay on the same side of the isocline, which guarantees stability, given the direction of selection in the areas surrounding it.

Appendix C. Why the simulation results differ (a bit) from the prediction of the simplified analytical model

In the analytical model, the direction of selection is determined for a population where all individuals have the same level of cooperativeness as well as the same grouping tendency. Also it is assumed that individual grouping tendencies translate to group size in a deterministic way. An equilibrium of the analytical model therefore represents a population that is monomorphic for cooperative and for grouping tendencies, and in which, since group formation is deterministic, groups of equal size form.

In the simulations, in contrast, cooperative and grouping tendencies are coded as diploid polygenic traits that naturally maintain some variance in the population. Also the group formation process implies that the group size an individual will face is not deterministic. Even if the population were monomorphic both for grouping and for cooperative tendencies, the group size that an individual faces would still be a draw from a random distribution over group sizes. The size and cooperativeness of the group an individual will find itself in therefore is a random variable that depends on an individual's own cooperative tendency and its grouping tendency, as well as on those of the others in the population.

For a precise analytical solution of the simulation model, one would need to look for a stable distribution of the two tendencies, where not only the means should not change, but also the actual distributions should be stable. As a first simplifying step, however, it seems natural to expect that a close match would be provided by looking only at whether an average individual could benefit from deviating from the average. This would ask for a solution of the following system:

$$\frac{d\mathbb{E}f_i}{d\gamma_i}\bigg|_{\gamma_i = \overline{\overline{\gamma}}} = 0 \quad \text{and} \quad \frac{d\mathbb{E}f_i}{d\delta_i}\bigg|_{\delta_i = \overline{\delta}} = 0$$

Here γ_i and δ_i are an individual's own cooperative and grouping tendency, respectively, and $\overline{\overline{\gamma}}$ and $\overline{\delta}$ are the average cooperative and grouping tendency in the population. The expectation is taken over the distribution of group sizes and compositions. This implies that for every value for the average cooperative and grouping tendencies, one would still need to find the correct distribution of cooperative and grouping tendencies over the population that goes with it. And with those at hand, one would then have to find the probability distribution over different group sizes and different average cooperative tendencies of those groups that these imply for the individual. Finding these distributions analytically is impossible (see the online tutorial at http://staff.feweb.vu.nl/j.garcia/sociality for a description of the group formation procedure in order to get an impression of the second step).

This motivates the choice for the following simplifying assumptions. We assume that the population is monomorphous with respect to grouping and cooperative tendencies, and we assume that group formation is deterministic. Then we solve the following system:

$$\frac{df_i}{d\gamma_i}\bigg|_{\gamma_i=\overline{\gamma}}=0 \quad \text{and} \quad \frac{df_i}{d\delta_i}\bigg|_{\delta_i=\overline{\delta}}=0$$

Here $\overline{\gamma}$ is the average cooperative tendency in group, which by the assumption of monomorphous populations also equals the average cooperative tendency $\overline{\overline{\gamma}}$ in the whole population.

Below we discuss both simplifying assumptions in reverse. First Appendix C.1 considers how the group formation process being random makes a difference. For our fitness functions, being in a group that is too small is worse than being in a group that is too large, which suggests that the isoclines for the simulations could be right of the isoclines in the monomorphic analytical model; it might be optimal to be on the safe side. Then Appendix C.3 considers how it matters that in the actual simulation model the population is not monomorphic. We do mostly get results where fitness is maximized at intermediate values for cooperative and grouping tendencies. In those cases sexual reproduction always maintains variance in the population. Because the fitness function is not symmetric around the optimum, we will see that the population being a balanced polymorphism amplifies the effect of the asymmetry in the fitness function described above. An intuitive way of rephrasing this is that individuals not only want to be on the safe side themselves, but they also want to be on the safe side as to what type their offspring will be.

Of course we would like to check if this indeed adds up to the difference between the simplified analytical model and the more complex simulation model. By lack of an analytical expression for the equilibrium distribution of types in the population and for the group size distributions given the distribution of types, we will have to rely for both on the observed distributions, once in equilibrium. These distributions will be different for different parameter values, so what we did in the appendix is just look at one choice for the parameter values, just consider group size, and see if these distributions agree with the explanation of the difference. It turns out that the agreement is very good.

C.1. Optimizing over probability distributions is not the same as finding the optimal (deterministic) group size

Individuals in the simulations are restricted to "choose" between different random variables. They cannot choose a fixed group size, but they do have a choice between different probability distributions over group sizes. That implies that a fixed point of the dynamics is not just a combination of values for cooperativeness and group size, but a *distribution* of group sizes and cooperative tendencies that remains stable. Below we will see that the mean of the best choice for a distribution over group sizes in general does not coincide with the optimum of the fitness function, which is where we would expect to find the population if individuals could choose a fixed, deterministic group size, rather than a distribution over group sizes.

We would like to use a few examples to show how both the set of probability distributions they can choose from and the shape of the fitness function can make the group size faced by the average individual differ from the optimum of the fitness function. The examples are a bit constructed, in order to keep the point simple, but at the end of this appendix, we will look at one equilibrium point from the simulations and see what the distributions are that individuals actually can choose from. For elegance and conciseness, we use continuous variables. Everything could be done with

discrete valued probability distributions just the same, but that would only be harder to digest.

We first assume that the number of offspring depends on group size s in the following way:

$$\begin{cases} s-1 & \text{if } s \in [1,1+\gamma] \\ 2\gamma-(s-1) & \text{if } s \in [1+\gamma,1+2\gamma] \\ 0 & \text{if } s \in [1+2\gamma,\infty] \end{cases}$$

which makes a tent-shaped function with the peak at $1+\gamma$ (Fig. C1).

Then suppose that the density function of the random variable of the size of a randomly chosen group is

$$l_t(s) = \frac{1}{\ln(t+1)}s^{-1}, \quad s \in [1,1+t]$$

With this random variable, expected group size is

$$\int_1^{1+t} s \cdot \frac{1}{\ln(1+t)}s^{-1}\,ds = \int_1^{1+t}\frac{1}{\ln(1+t)}\,ds = \frac{t}{\ln(1+t)}$$

More important, however, is that, if we take a random draw from the population, with every individual having equal probability, then this randomly drawn individual is s/u times as likely to come from a particular group of size s than it is to come from a particular group of size u. So the probability that a randomly drawn individual finds itself in a group of size s should be proportional to s times $l_t(s)$. This gives the uniform distribution on $[1,1+t]$, written as follows:

$$k_t(s) = \frac{1}{t}, \quad s \in [1,1+t]$$

The latter will be referred to as the group size faced by the average individual. This random variable has expected group size

$$\int_1^{1+t} s\frac{1}{t}\,ds = \frac{1}{t}\left[\frac{1}{2}s^2\right]_1^{1+t} = 1+\frac{1}{2}t$$

Both densities are pictured in Fig. C2.

For this combination of the fitness function and the set of distributions the individual can choose from, the expected fitness is

$$\begin{cases} \int_1^{1+t}\frac{1}{t}(s-1)\,ds & \text{if } t \in [0,\gamma] \\ \int_1^{1+\gamma}\frac{1}{t}(s-1)\,ds + \int_{1+\gamma}^{1+t}\frac{1}{t}(2\gamma-(s-1))\,ds & \text{if } t \in [\gamma,2\gamma] \\ \int_1^{1+\gamma}\frac{1}{t}(s-1)\,ds + \int_{1+\gamma}^{1+2\gamma}\frac{1}{t}(2\gamma-(s-1))\,ds & \text{if } t \in [2\gamma,\infty] \end{cases}$$

Below, this is plotted as a function of the group size faced by the average individual $(1+\frac{1}{2}t)$. The figure shows that for the optimal distribution, the group size faced by the average individual is
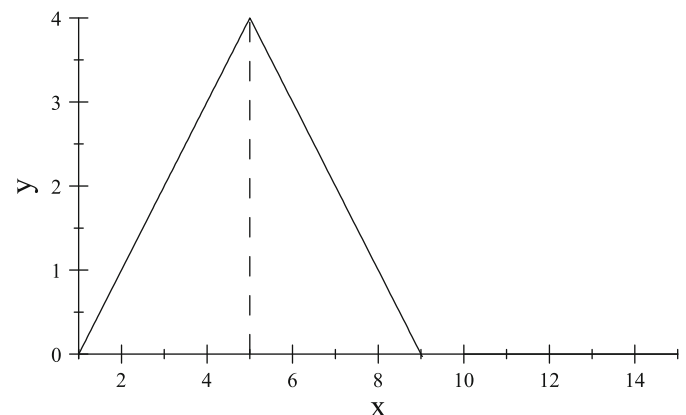
Fig. C1. Tent-shaped fitness function with $\gamma=4$.
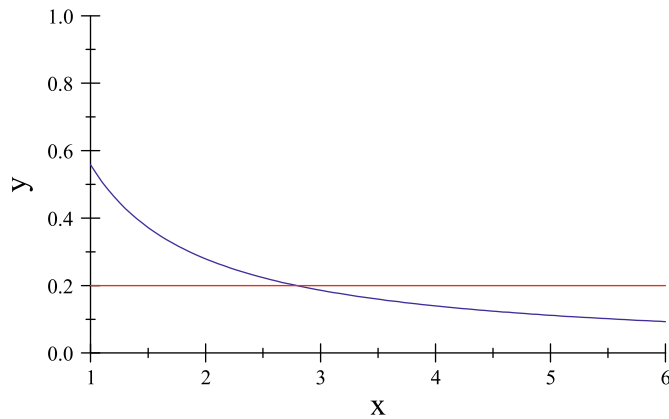
Fig. C2. Densities of group size (blue) and group size faced by a randomly chosen individual (red) for $t=5$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
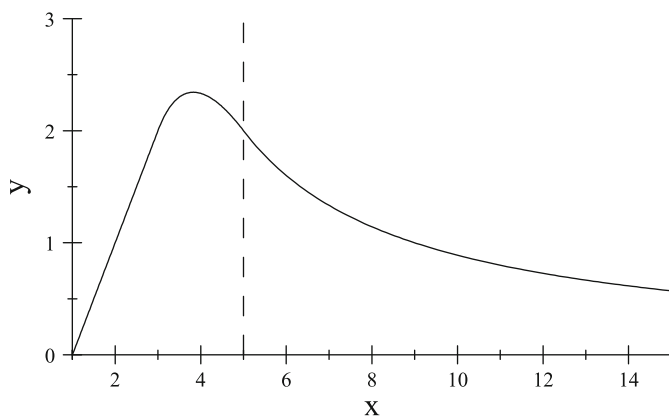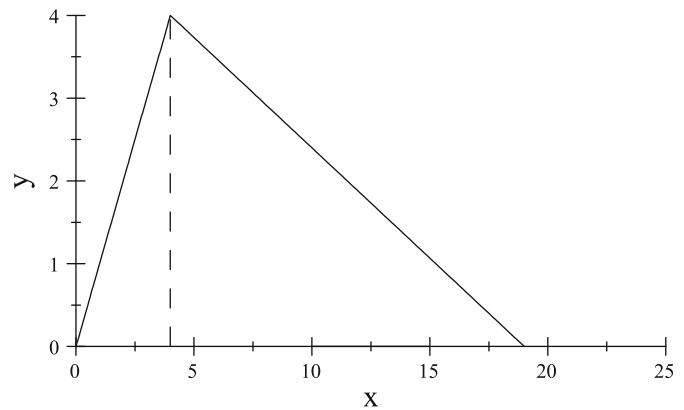

Fig. C4. Asymmetric fitness function with $\gamma=4, \delta=15$.

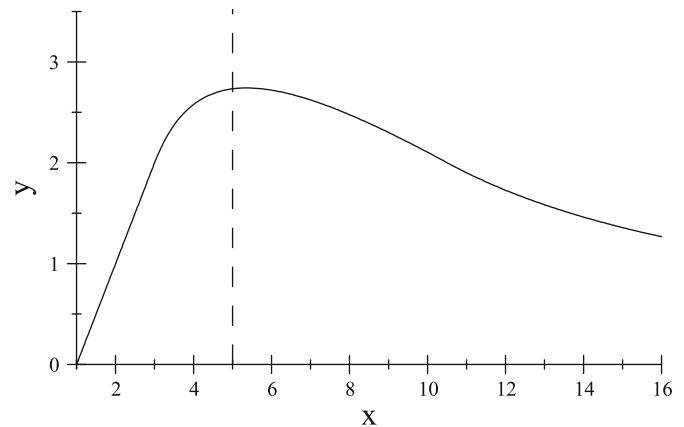**Fig. C3.** Expected payoff for $\gamma=4$.

Fig. C5. Expected fitness for $\gamma=4$ and $\delta=15$.

smaller than the size in the optimum of the fitness function. This is understandable, because the optimum lies in a point where the average of the fitness function over $[1,1+t]$ equals the value of that same function in the point $1+t$, which cannot be true if $t \geq 2\gamma$ (Fig. C3).

The actual optimum can be found by taking the derivative to t, which gives $t=\sqrt{2}\gamma$. Hence it holds here that for any of the tent-shaped fitness functions, that is, for any γ, the group size faced by the average individual is smaller than the optimum of the fitness function; $1+\frac{1}{2}t=1+\frac{1}{2}\sqrt{2}\gamma < 1+\gamma$.

An example of a set of distributions for which the mean of the optimal distribution coincides with the peak, when using this fitness function, is the set of uniform distributions on $[t-1,t+1]$, where the individual can choose a $t>2$, and an example where the mean of the optimal distribution lies right of the optimum of the fitness function, is the set of probability distributions where the individual finds itself in a group of size $1+t$ with probability $\frac{2}{3}$ and in a group of size $1+4t$ with probability $\frac{1}{3}$.

Now suppose we keep to the first set of probability distributions—uniform on $[1,1+t]$—but replace the fitness function with the following one:

$$\begin{cases} s-1 & \text{if } s \in [1,1+\gamma] \\ \frac{\gamma}{\delta}(\gamma+\delta-(s-1)) & \text{if } s \in [1+\gamma,1+\gamma+\delta] \\ 0 & \text{if } s \in [1+\gamma+\delta,\infty] \end{cases}$$

which is, unlike the first fitness function, not symmetric if $\gamma \neq \delta$ (Fig. C4):

Then expected fitness becomes

$$\begin{cases} \int_1^{1+t} \frac{1}{t}(s-1)\,ds & \text{if } t \in [0,\gamma] \\ \int_1^{1+\gamma} \frac{1}{t}(s-1)\,ds + \int_{1+\gamma}^{1+t} \frac{1}{t}\frac{\gamma}{\delta}(\gamma+\delta-(s-1))\,ds & \text{if } t \in [\gamma,\gamma+\delta] \\ \int_1^{1+\gamma} \frac{1}{t}(s-1)\,ds + \int_{1+\gamma}^{1+\gamma+\delta} \frac{1}{t}\frac{\gamma}{\delta}(\gamma+\delta-(s-1))\,ds & \text{if } t \in [\gamma+\delta,\infty] \end{cases}$$

The optimum can again be found by taking the derivative to t, which gives $t=\sqrt{\gamma^2+\delta\gamma}$. Therefore, if we choose $\delta>3\gamma$, we now get $1+\frac{1}{2}t=1+\frac{1}{2}\sqrt{\gamma^2+\delta\gamma}>1+\frac{1}{2}\sqrt{4\gamma^2}=1+\gamma$. This implies that for $\delta>3\gamma$, which for instance is the case in Fig. C5, the group size faced by the average individual in the optimal choice for a distribution is now larger than the optimum of the fitness function.

C.2. Data from the simulations

We will make a similar comparison for a simulation run; we will compare the optimal group size with the expected group size for the optimal probability distribution.

The functions f_i, g_i and h_i from the paper are all asymmetric in the same way that the last fitness function is; mistakes left from the optimum are worse than mistakes of the same size right of the optimum. For the simulation model, the distributions over group sizes that an individual faces, depending on the value of its grouping tendency, are actually rather hard to compute analytically. In the group formation process there, individuals leave a pool in random order, and one after the other seek admittance to a sequence of already formed groups. Their probability of being
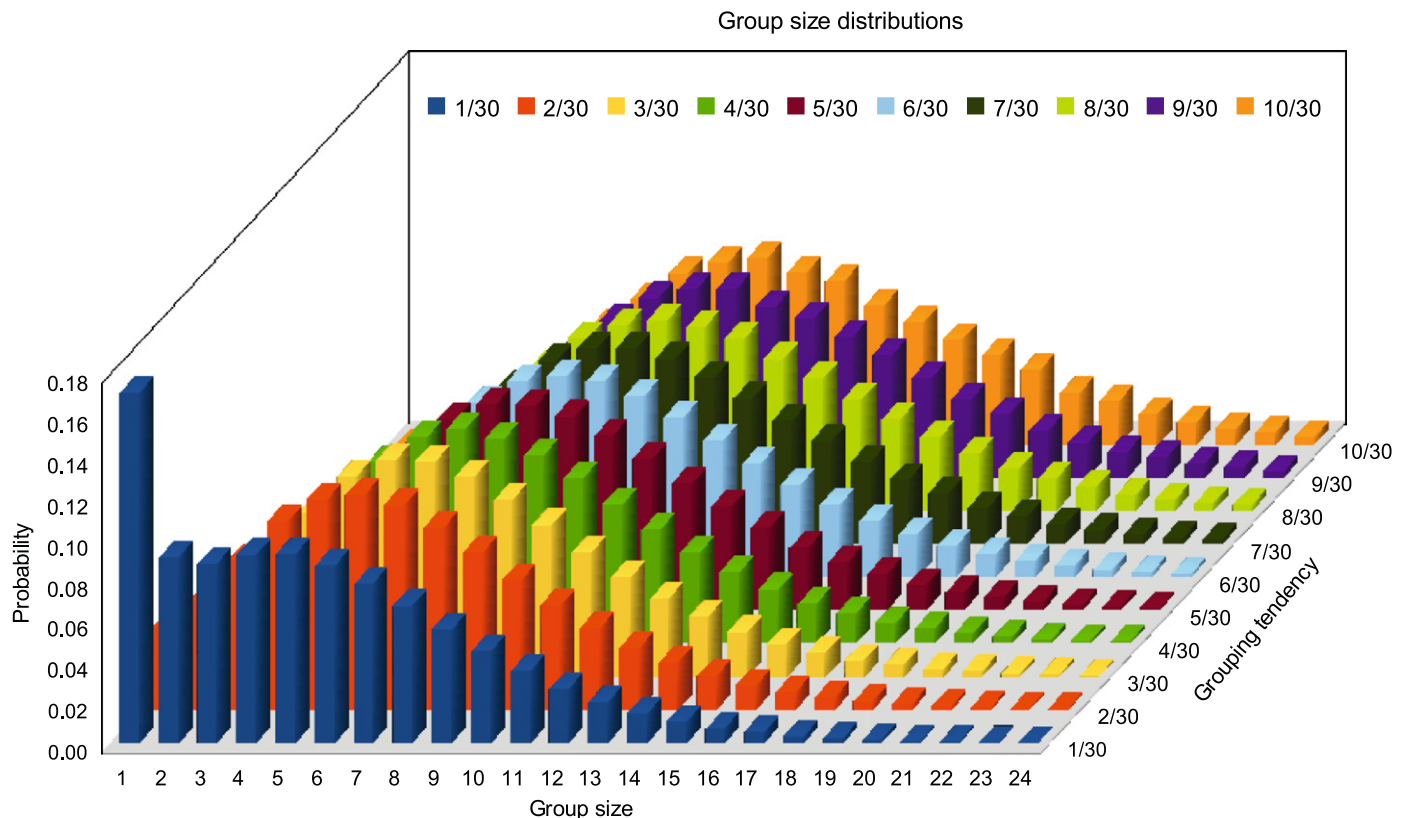
Fig. C6. Conditional distributions. For the function f_i with parameter values $\beta = 0.31$ and $c = \frac{1}{15}$, the simulation program counted how many times individuals with a given grouping tendency find themselves in groups of different sizes. This gives the empirical distributions pictured above. The distribution for grouping tendency 0 is not pictured, because they end up by themselves in groups of size 1 with probability 1, and grouping tendencies larger than $\frac{10}{30}$ occurred too infrequently to produce proper distributions.

admitted depends on their own grouping tendency and the average grouping tendency of the members of the group it seeks admittance to. This implies that the probabilities of events at every step of this process are not independent, which makes computing the probabilities for the group sizes that a randomly chosen individual faces, given its grouping tendency, seriously time consuming. Once we have reached an equilibrium in a simulation run, we can however look at the empirical distributions there. Fig. C6 pictures this empirical set of distributions, where every grouping tendency comes with a different distribution; the frequencies with which individuals with a low grouping tendency find themselves in groups of different sizes make the distributions up front, starting with the blue one, and the frequencies with which individuals with a high grouping tendency find themselves in groups of different sizes make the distribution at the back. Obviously, the distributions shift to the right as grouping tendencies increase; with a high grouping tendency, one can still end up in a small group, but the odds are just getting smaller as grouping tendency increases.

For each of those distributions, we computed the expected fitness and group size faced by the average individual (see Fig. C7). It turns out that expected fitness is maximized at a grouping tendency of $\frac{3}{30}$, which comes with a group size faced by the average individual of 8.2. This we should compare to what we get if we just optimize over group size. Optimizing over group size[2] gives a group size of 6.1. Note that it is not that an average

group size closer to 6.1 was not feasible with the different discrete options for grouping tendencies; with a grouping tendency of $\frac{1}{30}$ the average group size was 5.79, which is much closer to 6.1 than 8.2 is. Note that the value of the fitness function in the integer that is closest to the optimal group size (6) to is 3.77 and that expected fitnesses for all grouping tendencies in Fig. C7 are lower than that. This does make sense, because they are restricted to probability distributions, which imply that one cannot avoid being in groups of the suboptimal size; one can only balance the odds.

Still the actual group size faced by the average individual is not 8.2 but 8.8, and grouping tendencies of larger than $\frac{3}{30}$ constitute a substantial part of the population; average grouping tendency is $0.14 > \frac{3}{30}$. While choosing between distributions over group sizes—rather than between deterministic group sizes—explains part of the difference (the part between 6.1 and 8.2), the remaining part of the difference in group size can be explained by the fact that the equilibrium is (in fact, must be) a balanced polymorphism, where the expected fitness as a function of grouping tendency is asymmetric. Especially the low fitness for a grouping tendency of $\frac{0}{30}$ plays a role here (see Fig. C7).

C.3. Balanced polymorphism

The standard example of a balanced polymorphism is sickle-cell anaemia; in the presence of malaria threat, the heterozygote has the highest fitness, while both homozygotes have a lower fitness. Because the one homozygote that causes sickle-cell anaemia has a (much) lower fitness than the other homozygote

---

[2] In all the computations we used the average cooperativeness from the run to compute fitness.
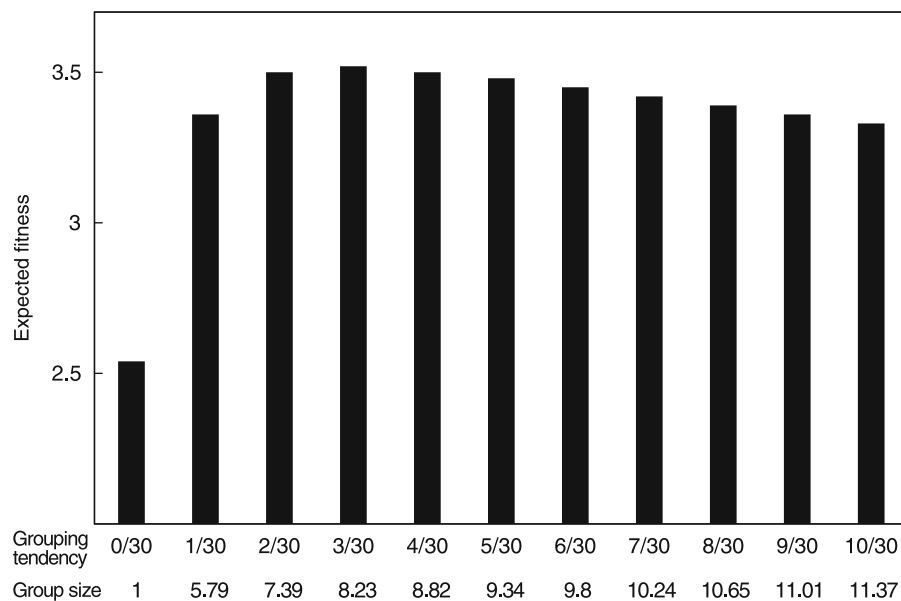
Fig. C7. Expected fitnesses. The average cooperativeness from the run with function f_i and parameter values $\beta = 0.31$ and $c = \frac{1}{15}$ was 0.406. With this value, the expected fitness is computed given the distributions as follows: $\mathbb{E}(f_i|\overline{\gamma} = 0.406) = \sum_{n=0}^{40} \mathbb{P}(\text{group size} = n) \cdot f_i(n, 0.406, 0.406)$.

that has an increased risk of contraction of malaria, the equilibrium value for the allele in the population is (well) below 50%, even though the heterozygote has maximum fitness and comes with 50% of each allele. One way to intuitively understand how this works is to see that selection is much harder on instances of having too much of the gene for malaria protection (which implies getting sickle-cell anaemia) than it is on having too little and suffering an increased risk of contracting malaria. Only in the case where both homozygotes have the same suboptimal fitness, does the frequency in the polymorphism coincide with the frequency in the optimal genotype.

The principle in this case is the same. It is driven by the fact that, just as with sickle-cell anaemia, the fitnesses around the optimum are not symmetric; if we look at Fig. C7, we see that individuals with grouping tendencies to the left of the optimum do worse than those to the right; $\frac{2}{30}$ does worse than $\frac{4}{30}$, $\frac{1}{30}$ does worse than $\frac{5}{30}$, and $\frac{0}{30}$ does much worse than $\frac{6}{30}$. Here selection disproportionally punishes having too little of the gene. The population average in equilibrium is therefore higher than the value that corresponds to the frequency in a population where all individuals are at the maximum expected fitness. The group size will therefore be even further to the right; 8.8 instead of 8.2. Especially the low fitness for grouping tendency at 0 plays a role here.

We also find that the balanced polymorphism part of the difference goes away if the version without sexual reproduction is run with a very low mutation rate. (The mutation rate needs to be low in order to be comparable, because mutation is much more biased against the common allele in the version without sex. It does take much longer to get to equilibrium though, with very low mutation rates.)

For grouping tendency, and with sexual reproduction, one could therefore say that the asymmetry of the fitness function affects the mean twice; once because there is uncertainty about the group size one will find itself in, given ones own grouping tendency, and once more because there in uncertainty about the grouping tendency of ones offspring. In both cases, uncertainty increases the mean, because deviations from the maximum of the fitness function to smaller groups are worse than deviations to larger groups.

References

Alexander, R.D., 1974. The evolution of social behavior. Annu. Rev. Ecol. Syst. 5, 325–383.

Avilés, L., 1997. Causes and consequences of cooperation and permanent-sociality in spiders. In: Choe, J., Crespi, B. (Eds.), Evolution of Social Behaviour in Insects and Arachnids. Cambridge University Press, Cambridge, pp. 476–498.

Avilés, L., 1999. Cooperation and non-linear dynamics: an ecological perspective on the evolution of sociality. Evol. Ecol. Res. 1, 459–477.

Avilés, L., 2002. Solving the freeloaders paradox: genetic associations and frequency dependent selection in the evolution of cooperation among relatives. Proc. Natl. Acad. Sci. 99, 14268–14273.

Avilés, L., Abbot, P., Cutter, A.D., 2002. Population ecology, nonlinear dynamics, and social evolution. I. Associations among nonrelatives. Am. Nat. 159, 115–127.

Avilés, L., Fletcher, J., Cutter, A.D., 2004. The kin composition of social groups: trading group size for degree of altruism. Am. Nat. 164, 132–144.

Avilés, L., Tufiño, P., 1998. Colony size and individual fitness in the social spider *Anelosimus eximius*. Am. Nat. 152, 403–418.

Berryman, A.A., 1973. Population-dynamics of fir engraver, Scolytus ventralis (Coleoptera-Scolytidae). Analysis of population behavior and survival from 1964 to 1971. Can. Entomol. 105, 1465–1488.

Booth, D.J., 1995. Juvenile groups in a coral-reef damselfish: density-dependent effects on individual fitness and population demography. Ecology 76, 91–106.

Buss, L.W., 1981. Group living, competition, and the evolution of cooperation in a sessile invertebrate. Science 213, 1012–1014.

Caraco, T., Wolf, L.L., 1975. Ecological determinants of groups sizes of foraging lions. Am. Nat. 109, 343–352.

Cash, K., McKee, M., Wrona, F., 1993. Short- and long-term consequences of grouping and group foraging in the free-living flatworm Dugesia tigrina. J. Anim. Ecol. 62, 529–535.

Emlen, S.T., 1991. Cooperative breeding in birds and mammals. In: Krebs, J.R., Davis, N.B. (Eds.), Behavioral Ecology: An Evolutionary Approach. Blackwell Scientific, Oxford, pp. 305–339.

Giraldeau, L.-A., 1988. The stable group and the determinants of foraging group size. In: Slobodchikoff, C.N. (Ed.), The Ecology of Social Behavior. Academic Press, San Diego, CA, pp. 33–53.

Giraldeau, L.-A., Caraco, T., 1993. Genetic relatedness and group size in an aggregation economy. Evol. Ecol. 7, 429–438.

Heinsohn, R.G., 1992. Cooperative enhancement of reproductive success in white-winged choughs. Evol. Ecol. 6, 97–114.

Higashi, M., Yamamura, N., 1993. What determines animal group size? Insider–outsider conflict and its resolution. Am. Nat. 142, 553–563.

Jarman, P.J., 1974. The social organisation of antelope in relation to their ecology. Behaviour 48, 215–268.

Jarman, P.J., 1982. Prospects for interspecific comparison in sociobiology. In: King's College Sociobiology Group, Current Problems in Sociobiology. Cambridge University Press, Cambridge, pp. 323–342.

Jarman, P.J., Jarman, M.V., 1979. The dynamics of ungulate social organization. In: Sinclair, A.R.E., Norton-Griffiths, M. (Eds.), Serengeti: Dynamics of an Ecosystem. Chicago University Press, Chicago, IL, pp. 185–200.

Jeanne, R.L., Nordheim, E.V., 1996. Productivity in a social wasp: per capita output increases with swarm size. Behav. Ecol. 7, 43–48.

Kerr, B., Godfrey-Smith, P., 2002. Individualist and multi-level perspectives on selection in structured populations. Biol. Philos. 17, 477–517.

Komdeur, J., 1994. Experimental evidence for helping and hindering by previous offspring in the cooperative-breeding Seychelles warbler *Acrocephalus sechellensis*. Behav. Ecol. Sociobiol. 34, 175–186.

Krause, J., Ruxton, G.D., 2002. Living in Groups. Oxford University Press.

Lehmann, L., Keller, L., 2006. The evolution of cooperation and altruism—a general framework and a classification of models. J. Evol. Biol. 19, 1365–1376.

Nudds, T.D., 1978. Convergence of group size strategies by mammalian social carnivores. Am. Nat. 112, 957–960.

Pepper, J.W., 2000. Relatedness in group-structured models of social evolution. J. Theor. Biol. 206 (3), 355–368.

Pulliam, H.R., Caraco, T., 1984. Living in groups: Is there an optimal group size?. In: Krebs J.R., Davis, N.B. (Eds.), Behavioral Ecology: An Evolutionary Approach. Blackwell Scientific, Oxford, pp. 122–147.

Raffa, K.F., Berryman, A.A., 1987. Interacting selective pressures in conifer-bark beetle systems: a basis for reciprocal adaptations? Am. Nat. 129 234–262.

Reiczigel, J., Lang, Z., Rózsa, L., Tóthmérész, B., 2008. Measures of sociality: two different views of group size. Anim. Behav. 75 (2), 715–721.

Slobodchikoff, C.N., 1984. Resources and the evolution of social behavior. In: Price, P.W., Slobodchikoff, C.M., Gaud, W.S. (Eds.), A New Ecology: Novel Approaches to Interactive Systems. Wiley, New York, pp. 227–251.

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 (3), 589–600.

Vehrencamp, S.L., 1983. A model for the evolution of despotic versus egalitarian societies. Anim. Behav. 31, 667–682.

Wiklund, C.G., Andersson, M., 1994. Natural selection of colony size in a passerine bird. J. Anim. Ecol. 63, 765–774.

Wilson, D.S., 1979. Structured demes and train-group variation. Am. Nat. 113, 606–610.

Wilson, D.S., 2004. What is wrong with absolute individual fitness? Trends in Ecology and Evolution 19, 245–248.